Stochastic Calculus and Mathematical Finance

2012

P. Ouwehand

Department of Mathematical Sciences
Stellenbosch University

Contents

Chapter 1

Brownian Motion

1.1 Modelling Stock Returns

Any model of stock price behaviour must be *stochastic*, i.e. it must incorporate the random nature of price behaviour. The simplest such models are *random walks*: Let $X_t, t = 1, 2, \ldots$ be a family of distributed random variables, and let S_0 be the stock price at $t = 0$. We might (naively) attempt to model the stock price process by

$$S_t = S_{t-1} + X_t \quad \text{i.e.} \quad S_t = S_0 + \sum_{u=1}^{t} X_u$$

The intuition behind this is that the price at time t equals the price at time $t-1$ plus a "random shock", modelled by X_t.

We should also assume that these shocks are *independent*. Why? If we could predict today that the stock price is going to go up *tomorrow*, this makes the stock more attractive today. Thus more people would buy it today, forcing the stock price up *today*, until it reaches the level predicted. Thus any change in the stock price must essentially be unpredictable. This is just a version of *Efficient Markets Hypothesis*, which, loosely, asserts that all available information about a corporation is instantly reflected in its stock price. Thus future changes in price are not dependent on past changes in price.

There are several reasons why a random walk model of stock prices is inadequate, but an obvious one is that it doesn't take into account scale. for stock prices, we expect the change in price to be *proportional* to the current price. To see this, consider two companies in two parallel universes, A and B. The universes and the companies are identical, except for one thing. In universe A, the company has issued 100 shares, each trading at \$100. In universe B, the company has undertaken a 2–for–1 stock split, so that it has issued 200 shares, each trading at \$50. Both companies are otherwise identical, e.g. they are both worth \$10 000. One day an earthquake cause massive damage, and both companies lose half their value. The shares in universe A now trade at \$50, whereas those in universe B trade at \$25. Thus the share price has not dropped by the same amount in both universes: Each share has lost the same *proportion* of its value.

Simply put, if investors require a return of 14%, then they require that return irrespective of whether the share price is \$50 or \$100.

The shares of A, B change by the same *factor*, i.e. they have exactly the same change in returns (but not the same absolute change in price). This is reflected in, e.g., the binomial

model, where shares can go up by a *factor* of u or down by a *factor* of $\frac{1}{u}$. But a multiplicative change in the stock price amounts to an additive change in the logarithm of the stock price:

$$S_{t+\Delta t} = u^{\pm 1} S_t \qquad \text{implies} \qquad \ln S_{t+\Delta t} = \ln S_t \pm \ln u$$

i.e. if we define the returns process R_t by $S_t = S_0 e^{R_t}$ (i.e. $R_t := \ln \frac{S_t}{S_0}$), and define $\delta := \ln u$, we have i.e.

$$R_{t+\Delta t} = R_t \pm \delta$$

A better random model of stock prices is therefore one in which the *returns* process R_t follows a random walk.

We now seek a continuous-time version of the random walk — a stochastic process that is changing because of random shocks at every instant in time. Consider a time interval $[0, T]$ and let N be a (large) integer. Define $\Delta t := \frac{T}{N}$. Let $X_n, n = 1, 2, 3, \ldots$ be independent Bernoulli random variables with

$$\mathbb{P}(X_n = \Delta x) = p \qquad \text{and} \qquad \mathbb{P}(X_n = -\Delta x) = 1 - p =: q$$

where $\Delta x > 0$. For $t = 0, \Delta t, 2\Delta t, \ldots, N\Delta t = T$, let $R_t := \sum_{k=1}^{n} X_k$, where $t = n\Delta t$. Thus R_t is a random walk, and

$$R_{t+\Delta t} = R_t \pm \Delta x$$

Some simple calculations yield

$$\mathbb{E}[R_t] = n(p-q)\Delta x = (p-q)\frac{\Delta x}{\Delta t} t \qquad \mathrm{Var}(R_t) = n(\Delta x^2 - (p-q)^2 \Delta x^2) = 4pq\frac{\Delta x^2}{\Delta t} t$$

Now suppose we can observe the process R_t and want $\mathbb{E}[R_t] = \mu t$ and $\mathrm{Var}(R_t) = \sigma^2 t$, where μ, σ are constants, and $\sigma > 0$. (We want $\sigma^2 > 0$, otherwise $\mathrm{Var}(R_t) = 0$, in which case R_t would be non–random.)

In the continuous limit, i.e. as $N \to \infty$ and $\Delta t \to 0$, we must have

$$(p-q)\frac{\Delta x}{\Delta t} \to \mu \qquad 4pq\frac{\Delta x^2}{\Delta t} \to \sigma^2$$

The first equation yield $\Delta x \approx \frac{\mu \Delta t}{p-q}$ when Δt is small. Substituting into the second equation, we see that

$$\frac{4pq}{(p-q)^2}\Delta t \approx \frac{\sigma^2}{\mu^2}$$

when Δt is small. Now since, $\Delta t \to 0$, we must have $\frac{4pq}{(p-q)^2} \to \infty$, for otherwise the product $\frac{4pq}{(p-q)^2}\Delta t$ would tend to 0, not $\frac{\sigma^2}{\mu^2}$. It is therefore necessary that $p - q \to 0$, and thus p, q must both tend to $\frac{1}{2}$ as $\Delta t \to 0$. From the fact that $4pq\frac{\Delta x^2}{\Delta t} \to \sigma^2$, we then see that we must have

$$\Delta x \approx \sigma\sqrt{\Delta t}$$

for small Δt.

We had $\Delta x \approx \frac{\mu \Delta t}{p-q}$ for small Δt, and thus $p - q \approx \frac{\mu}{\sigma}\sqrt{\Delta t}$. Since $p + q = 1$, we must have

$$p = \tfrac{1}{2}(1 + \tfrac{\mu}{\sigma}\sqrt{\Delta t}) \qquad q = \tfrac{1}{2}(1 - \tfrac{\mu}{\sigma}\sqrt{\Delta t})$$

As a check, note that

$$\mathbb{E}[R_t] = (p-q)\frac{\Delta x}{\Delta t}t = \frac{\mu}{\sigma}\sqrt{\Delta t}\frac{\sigma\sqrt{\Delta t}}{\Delta t}t = \mu t$$

and

$$\mathrm{Var}(R_t) = 4pq\frac{\Delta x^2}{\Delta t}t = (1 - \tfrac{\mu^2}{\sigma^2}\Delta t)\frac{\sigma^2\Delta t}{\Delta t}t = \sigma^2 t - \mu^2 t\Delta t \to \sigma^2 t$$

as should be the case.

We now have an idea of how to create a a continuous–time stochastic process R_t as the ($\Delta t \to 0$)–limit of a random walk. But the limit process has some peculiar features. For example

$$\Delta R_t \approx \pm\sigma\sqrt{\Delta t} \quad \text{is of the order of } \sqrt{\Delta t}$$

If $f(t)$ is a differentiable function, then

$$\Delta f(t) \approx f'(t)\Delta t \quad \text{is of the order of } \Delta t$$

Now when Δt is small, we see that $\sqrt{\Delta t}$ is much larger than Δt (Take, e.g. $\Delta t = 10^{-2n}$ and note that $\sqrt{\Delta t} = 10^{-n} = 10^n\Delta t$.) It follows that R_t cannot be differentiable as a function of t.

The probabilist will immediately want to know the distribution of R_t. Let $u(t,x)$ be the density of the random variable R_t, i.e.

$$u(t,x)\Delta x \approx \mathbb{P}(R_t \in [x, x+\Delta x])$$

At time $t+\Delta t$ the random walk can reach the point x in two ways: It can move right from the point $x-\Delta x$ at time t, with probability p, or it can move left from the point $x+\Delta x$, with probability q. Thus

$$u(t+\Delta t, x) = pu(t, x-\Delta x) + qu(t, x+\Delta x)$$

Now we Taylor expand up to order Δt. Firstly

$$u(t+\Delta t, x) \approx u(t,x) + u_t(t,x)\Delta t + o(\Delta t)$$

Next,

$$u(t, x\pm\Delta x) = u(t,x) \pm u_x(t,x)\Delta x + \tfrac{1}{2}u_{xx}(t,x)\Delta x^2 + o(\Delta x^2)$$

Here, we have taken a second–order Taylor expansion, because Δx is of the order $\sqrt{\Delta t}$, and Δx^2 of the order Δt. Putting these together, we obtain (at the point (t,x)):

$$u + u_t\Delta t = (p+q)u + (-p+q)u_x\Delta x + \tfrac{1}{2}(p+q)u_{xx}\Delta x^2$$

However, we know that $p = \frac{1}{2}(1+\frac{\mu}{\sigma}\sqrt{t})$ and that $\Delta x \approx \sigma\sqrt{\Delta t}$ and $p, q \to \frac{1}{2}$. Hence

$$u_t\Delta t = -(\tfrac{\mu}{\sigma}\sqrt{\Delta t})u_x(\sigma\sqrt{\Delta t}) + \tfrac{1}{2}u_{xx}\sigma^2\Delta t$$

which yields the following partial differential equation for the density of R_t.

$$u_t = -\mu u_x + \tfrac{1}{2}\sigma^2 u_{xx}$$

However, the PDE is not sufficient to determine the density u: It has many solutions. We seek a solution which has the following properties:

- For each $t \geq 0$, we have $\int_{-\infty}^{\infty} u(t,x)\, dx = 1$, because $u(t,x)$ is a density, and
- $u(0,x)$ is rather odd: We have $R_0 = 0$, and so

$$f(0) = \mathbb{E}[f(R_0)] = \int_{-\infty}^{\infty} f(x)u(0,x)\, dx$$

i.e. $u(0,x)$ is a "function" with the property that $\int_{-\infty}^{\infty} f(x)\, dx = f(0)$ for every function f. The "function" with this property is called the *Dirac delta* δ_0. It is not a function at all (but the simplest example of a so–called *generalized function* or *distribution* (in the sense of Schwartz).) Nevertheless, we can get some intuition as to how u ought to behave. We see that for t close to 0, the density $u(t,x)$ must be very small for $x \neq 0$, because R_t must be close to x when t is near zero. Yet the area under the curve is 1, i.e. $u(t,x)$ must be extremely peaked at around $x = 0$ and then rapidly drop off. We may thus think off $u(0,x) = \delta_0$ as a "function" which has

$$\delta_0(x) = 0 \text{ when } x \neq 0 \qquad \delta_0(0) = +\infty \text{ in such a way that } \int_{-\infty}^{\infty} \delta_0(x)\, dx = 1$$

Oddly enough, we can find such a function. The PDE for the density, derived by Einstein in 1905, is a version of the heat equation, derived by Fourier, which governs heat transfer. So this PDE was not new: It had been intensively studied by physicists, with $u(t,x)$ playing the role of the temperature at time t at a point x in an infinitely long rod. The *fundamental solution* or *Green's function* of such a PDE was well–known

$$u(t,x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}}$$

You can verify by direct differentiation that this function does, in fact, satisfy the PDE. You will also immediately recognize it as the density of an $N(\mu t, \sigma^2 t)$–random variable. Furthermore, for t near 0, such a random variable has very small standard deviation, and thus the density is extremely peaked around 0, just as we require.

It follows, therefore, that the density of R_t is $N(\mu t, \sigma^2 t)$. Of course, the Central Limit Theorem states that, subject to a moment condition, large sums of i.i.d. are roughly normally distributed, so we are not surprised. But here, we have in essence given a proof of the Central Limit Theorem by PDE methods, at least for random walks of the type described.

When we take $\mu = 0$ and $\sigma = 1$, we obtain one of the basic building blocks of financial modelling:

Definition 1.1.1 Standard *Brownian motion* is a continuous–time stochastic process $B_t, t \geq 0$ with the following properties:

(1) Each change

$$\begin{aligned} B_t - B_s = &(B_{s+h} - B_s) + (B_{s+2h} - B_{s+h}) \\ &+ \cdots + (B_t - B_{t-h}) \end{aligned}$$

is *normally distributed* with mean 0 and variance $t - s$.

(2) Each change $B_t - B_s$ is **independent** of all the previous values $B_u, u \leq s$.

(3) Each sample path $B_t, t \geq 0$ is (a.s.) **continuous**, and has $B_0 = 0$.

Now put

$$R_t = \mu t + \sigma B_t \qquad \text{so that} \qquad S_t = S_0 e^{\mu t + \sigma B_t}$$

It then follows easily that

$$R_t \sim N(\mu t, \sigma^2 t)$$

i.e. the standard Brownian motion can also be used to to model returns processes where $\mu \neq 0$ and $\sigma \neq 1$. The process R_t is called an *arithmetic Brownian motion* with drift rate μ and variance rate σ^2. We will also refer to σ as the *volatility.* The process S_t is a *geometric Brownian motion*, and is lognormally distributed (i.e. each $\log S_t$ is normally distributed).

1.2 Definition and Existence of Brownian Motion

A Brief History:

- Brown 1828 — pollen suspended in water jiggles about as if alive.
- Bachelier 1900 — Théorie de la Spéculation.
- Einstein 1905 — Evidence for existence of atoms.
- Wiener 1923 — Existence of Brownian motion as mathematical object.

Definition 1.2.1 A 1–dimensional stochastic process $(B_t : t \geq 0)$ is called a *standard Brownian motion* or a *Wiener process* if and only if

(i) *Continous sample paths:* The map $t \mapsto B_t(\omega)$ is continuous for almost all ω.

(ii) *Independent increments:* Given $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables

$$B_{t_k} - B_{t_{k-1}} \qquad k = 1, \ldots, n$$

are mutually independent.

(iii) *Normally distributed increments* If $0 \leq s < t$, then

$$B_t - B_s \sim N(0, t-s)$$

(iv) (Not essential) $B_0 = 0$ a.s.
If we have $B_0 = x$, then we refer to $(B_t)_t$ as a standard Brownian motion starting at x.

□

Condition (iv) is convenient to develop the theory. In practice, we may want a Brownian motion to start at some other point x (or even at a random variable B_0). Once you've understood some of the properties of Brownian motion, you will realize that this doesn't harm the theory in any way.

If you peruse some of the results that follow, you will perceive that Brownian motion has very, very strange properties indeed. You may end up doubting whether such a creature can possibly exist. Thus, without further ado:

Theorem 1.2.2 (Wiener)

Brownian motion exists

Let the sample space Ω be $\mathcal{C}[0,\infty)$, the set of all continuous functions from $[0,\infty)$ to $\mathbb{R}$. Let the $\mathcal{F}$ be the σ–algebra on Ω which is generated by the projection mappings $\pi_t : C[0,\infty) \to \mathbb{R}$, which are given by

$$\pi_t : f \mapsto f(t)$$

Then for every $x \in \mathbb{R}$ there exists a probability measure $\mathbb{P}^x$ on $(\Omega, \mathcal{F})$ such that $(\pi_t)_t$ is a Brownian motion starting at x.

$\dashv$

Why did we fix the sample space to be $\Omega = \mathcal{C}[0,\infty)$ in the theorem above? Brownian motion has a natural home, i.e. there is a *canonical* probability space on which Brownian motion is defined. As above, let $\mathcal{F} = \sigma(\pi_t : t \geq 0)$ be the smallest σ–algebra on $C[0,\infty)$ which makes every projection measurable. Then $(\pi_t)_{t\geq 0}$ is a sequence of $\mathbb{R}^+$–indexed random variables, and can therefore be thought of as a stochastic process on the measurable space $(C[0,\infty), \mathcal{F})$.

We now show the following remarkable fact: Given that a Brownian motion $(\hat{B}_t)_t$ (starting at 0) exists on *some* probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, we can define, for each $x \in \mathbb{R}$, a measure $\mathbb{P}^x$ on $(C[0,\infty), \mathcal{F})$ such that $(\pi_t)_t$ is a Brownian motion on $(C[0,\infty), \mathcal{F}, \mathbb{P}^x)$, but starting at x.

Here's how we do this: First note that we can regard $\hat{B}$ as a function from $\hat{\Omega}$ to $C[0,\infty)$. For each $\hat{\omega} \in \hat{\Omega}$, $\hat{B}(\hat{\omega})$ can be thought of as a continuous function whose value at t is $\hat{B}_t(\hat{\omega})$. Thus

$$\hat{B} : \hat{\Omega} \to C[0,\infty) \qquad \hat{B}(\hat{\omega}) : t \mapsto \hat{B}_t(\hat{\omega})$$

(If necessary, redefine $\hat{B}_t(\hat{\omega})$ to be zero on the nullset of those $\hat{\omega}$ for which the sample path $(\hat{B}_t(\hat{\omega}))_t$ is not continuous). Now if $A \in \mathcal{F}$, then $A \subseteq C[0,\infty)$, i.e. A is a set of continuous functions. Define the measure $\mathbb{P}^0$ on $\mathcal{F}$ by

$$\mathbb{P}^0(A) = \hat{\mathbb{P}}(\{\hat{\omega} \in \hat{\Omega} : \hat{B}(\hat{\omega}) \in A\})$$

i.e. $\mathbb{P}^0 = \hat{\mathbb{P}} \circ \hat{B}^{-1}$, so that $\mathbb{P}^0$ is essentially the distribution of $\hat{B}$ under $\hat{\mathbb{P}}$. In particular, if $C \in \mathcal{B}(\mathbb{R})$, then

$$\mathbb{P}^0(\pi_t \in C) = \hat{\mathbb{P}}(\hat{B} \in \pi_t^{-1}(C)) = \hat{\mathbb{P}}(\pi_t(\hat{B}) \in C) = \hat{\mathbb{P}}(\hat{B}_t \in C)$$

It follows that π_t has the same distribution under $\mathbb{P}^0$ as does $\hat{B}_t$ under $\hat{\mathbb{P}}$. We now show that $(\pi_t)_t$ is a Brownian motion on $(C[0,\infty), \mathcal{F}, \mathbb{P}^0)$, i.e. we check (i)–(iv) of Definition 1.2.1. Since the sample space consists of continuous functions, we will write f for a sample point, rather than ω.

(i) For a sample point f, the map $t \mapsto \pi_t(f)$ is just the map f (because $\pi_t(f) = f(t)$), and thus certainly continuous. Thus *every* sample path is continuous, and not merely almost every sample path.

(ii) Since $\pi_t - \pi_s$ has the same distribution under $\mathbb{P}^0$ as does $\hat{B}_t - \hat{B}_s$ under $\hat{\mathbb{P}}$, it follows that $\pi_t - \pi_s$ is $N(0, t-s)$ under $\mathbb{P}^0$.

(iii) Just like (ii).

(iv) $\mathbb{P}^0(\pi_0 = 0) = \hat{\mathbb{P}}(\hat{B}_0 = 0) = 1$, i.e. $\pi_0 = x$ almost surely.

The measure $\mathbb{P}^0$ is called *Wiener measure*. It can be thought of as a measure on the set of all sample paths, and this provides a very intuitive way of looking at things. For example, since $\pi_0 = 0$ a.s., the set of continuous functions f with $f(0) \neq 0$ is a $\mathbb{P}^0$– nullset. Similarly, we shall see later, the set of differentiable paths is a nullset, because a Brownian sample path is nowhere differentiable almost surely.

Now let $x \in \mathbb{R}$, and define $\mathbb{P}^x$ on $(\mathcal{C}[0,\infty), \sigma(\pi_t : t \geq 0))$ by

$$\mathbb{P}^x(F) = \mathbb{P}^0(F - x)$$

Here F is a set of continuous functions, and $F - x = \{f - x : f \in F\}$ is another set of continuous functions. It is easy to see that $F - x$ is measurable if F is, so that this definition makes sense. Note now that

$$\mathbb{P}^x(\pi_0 = x) = \mathbb{P}^x(\{f : f(0) = x\}) = \mathbb{P}^0(\{f - x : f(0) = x\}) = \mathbb{P}^0(\{g : g(0) = 0\} = 1$$

It is now easy to see that $(\pi_t)_t$ is a Brownian motion starting at x.

Thus the same process is a Brownian motion under each $\mathbb{P}^x$, but it starts at a different place under each measure!

Remarks 1.2.3 We have two candidates for "natural" σ–algebra on $C[0,\infty)$. The first is $\mathcal{F} = \sigma(\pi_t : t \geq 0)$, the σ–algebra generated by the projections, used above. The second is obtained via topological considerations. We can define a metric ρ on $C[0,\infty)$ as follows: First define ρ_n on $C[0,\infty)$ by

$$\rho_n(f,g) = \sup_{0\leq t\leq n} |f(t) - g(t)|$$

Then define

$$\rho(f,g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_n(f,g)}{1+\rho_n(f,g)}$$

Then ρ is a metric on $C[0,\infty)$, and thus induces a topology. This topology is called the *topology of uniform convergence on compact sets*, because $f_n \to f$ in this topology if and only if $f_n \to f$ uniformly on every compact set.

Now any topological space X has a natural σ–algebra, namely its Borel algebra. This is the smallest σ–algebra on X which contains every open set. In our case, the Borel algebra $\mathcal{B}$ of the topological space $C[0,\infty)$ coincides with the algebra $\mathcal{F}$, i.e. our two "natural" σ–algebras coincide.

To see that this is so is an exercise. First show that each $\pi_t : C[0,\infty) \to \mathbb{R}$ is continuous (i.e. that if $\rho(f,g)$ is small, then $|f(t) - g(t)|$ is small as well). This shows that $\mathcal{F} \subseteq \mathcal{B}$. Next suppose that $f_0 \in C[0,\infty)$. Note that $\rho_n(f_0, f) = \sup_{\mathbb{Q}\cap[0,n]} |f_0(q) - f(t)|$ and conclude that ρ_n, hence ρ, is $\mathcal{F}$–measurable. Further note that $C[0,\infty)$ is separable, and thus has a countable dense set $\{f_n : n \in \mathbb{N}\}$. Finally, let F be a closed subset of $C[0,\infty)$, and note that $F = \{f \in C[0,\infty) : \inf_n \rho(f, f_n) = 0\}$. This proves that $\mathcal{F}$ contains every closed subset of $C[0,\infty)$.

□

We now say a few words about the measures $\mathbb{P}^x$. Let

$$p_t(a,b) = \frac{1}{\sqrt{2\pi t}}\, e^{\frac{-(b-a)^2}{2t}}$$

be the density function of a $N(0,t)$–random variable. If $A_1 \in \mathcal{B}(\mathbb{R})$, then clearly

$$\mathbb{P}^x(B_{t_1} \in A_1) = \int_{A_1} p_{t_1}(x,y)\, dy$$

We see that $p_{t_1}(x,y)\, dy$ can be thought of as the probability that B_t moves from point x at time $t = 0$ to point y at time $t = t_1$. The probability of ending up in the set A_1 is then given by summing over all $y \in A_1$, yielding the above integral.

Next, since B_{t_1} and $B_{t_2} - B_{t_1}$ are independent when $t_2 > t_1$, we have

$$\mathbb{P}^x(B_{t_1} \in A_1, B_{t_2} \in A_2) = \int_{A_1}\int_{A_2} p_{t_1}(x,x_1) p_{t_2-t_1}(x_1,x_2)\, dx_2\, dx_1$$

We can interpret $p_{t_1}(x,x_1)p_{t_2-t_1}(x_1,x_2)\, dx_1\, dx_2$ as the probability that B_t moves from point x at time $t = 0$ to point x_2 at time $t = t_2$, via point x_1 at time t_1. For this reason, the function $p_t(x,y)$ is known as the *transition density*: It governs how Brownian motion moves from point to point.

This generalizes: Let $0 = t_0 < t_1 < \cdots < t_n$, and also let $x_0 = x$ (the starting point). Then the *joint density function* of $(B_{t_0}, \ldots, B_{t_n})$ is given by

$$\mathbb{P}^x(B_{t_0} = x_0, \ldots, B_{t_n} = x_n) = \prod_{k=1}^{n} p_{t_k - t_{k-1}}(x_{k-1}, x_k)\, dx_1\, \ldots\, dx_n$$

so that, for example,

$$\mathbb{P}^x(B_{t_0} \in A_0, \ldots, B_{t_n} \in A_n) = I_{A_0}(x) \int_{A_1} \cdots \int_{A_n} \prod_{k=1}^{n} p_{t_k - t_{k-1}}(x_{k-1}, x_k)\, dx_n \ \ldots\ dx_1 \qquad (*)$$

Note that the σ–algebra $\mathcal{F} = \sigma(\pi_t : t \geq 0)$ on Wiener space is generated by the π-system of sets of the form $\{B_{t_0} \in A_0, \ldots, B_{t_n} \in A_n\}$. Thus $(*)$ shows how $\mathbb{P}^x$ is defined on a π–system generating $\mathcal{F}$. Kolmogorov's Extension Theorem and Kolmogorov's Continuity Theorem allow us to extend $(*)$ to prove the existence of a Brownian motion on $\mathcal{C}[0, \infty)$.

We can also generalize the definition of the measures $\mathbb{P}^x$: Let μ be a probability distribution, i.e. a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Define a measure $\mathbb{P}^\mu$ on $(\mathcal{C}[0,\infty), \sigma(\pi_t : t \geq 0))$ by

$$\mathbb{P}^\mu(F) = \int \mathbb{P}^x(F)\, d\mu$$

Then $(\pi_t)_t$ is a Brownian motion with the property that π_0 is μ–distributed. To see this, first note that, for each F, the map

$$x \mapsto \mathbb{P}^x(F)$$

is Borel measurable (i.e. measurable from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$): Certainly, if F is of the form $F = \{\pi_{t_0} \in A_0, \ldots, \pi_{t_n} \in A_n\}$, where $0 = t_0 < t_1 < \cdots < t_n$, we see that

$$\mathbb{P}^x(F) = I_{A_0}(x) \int_{A_1} \cdots \int_{A_n} p(t_1; x, y_1) \ldots p(t_n - t_{n-1}; y_{n-1}, y_n)\, dy_1 \ldots\ dy_n$$

so that $x \mapsto \mathbb{P}^x(F)$ is measurable. This proves that $x \mapsto \mathbb{P}^x(F)$ is measurable for all F in a π–system which generates $\sigma(\pi_t : t \geq 0)$. It now follows easily from Dynkin's Lemma that the map is measurable for all F.

Next, note that

$$\mathbb{P}^\mu(\pi_0 \in A) = \int \mathbb{P}^x(\pi_0 \in A)\, d\mu = \int I_A(x)\, d\mu = \mu(A)$$

This proves that π_0 is μ–distributed under $\mathbb{P}^\mu$. Now it is clear that $(\pi_t)_t$–sample paths are continuous. Furthermore, $\mathbb{P}^\mu(\pi_t - \pi_s \in A) = \int \mathbb{P}^x(\pi_t - \pi_s \in A)\, d\mu = \int \mu_Z(A)\, d\mu = \mu_Z(A)$, where Z is $N(0, t-s)$, and μ_Z is its distribution. It follows that $\pi_t - \pi_s$ has the distribution (under $\mathbb{P}^\mu$) of a $N(0, t-s)$–variable. Similarly, it is easily shown that the process $(\pi_t)_t$ is Gaussian under $\mathbb{P}^\mu$. We now invoke a future result, Proposition 1.5.2. According to this proposition, it only remains for us to show that $\mathrm{Cov}(\pi_t, \pi_s) = s$ (when $s \leq t$). But this is obvious, as it is so under each $\mathbb{P}^x$.

Very often, we will work with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a filtration $\mathcal{F}_t$, such that each B_t is $\mathcal{F}_t$–measurable, and such that instead of *(ii)* the stronger condition

$$\text{(ii)}' \qquad B_t - B_s \text{ is independent of } \mathcal{F}_s \text{ for all } 0 \leq s \leq t$$

Such a Brownian motion is called an $\mathcal{F}_t$–Brownian motion.

Definition 1.2.4 A stochastic process X_t is said to be a *stationary* process, or to have *stationary increments*, if and only if for any $0 \leq s < t$ and any $h > 0$, the random variables $X_t - X_s$ and $X_{t+h} - X_{s+h}$ are identically distributed.

□

It is easy to see directly from the definition that Brownian motion is a stationary process. Moreover, the increments over disjoint time intervals are independent.

Proposition 1.2.5 *If B_t is a standard Brownian motion, then*

a) The processes B_t, $B_t^2 - t$, and $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$ are martingales (with respect to the natural filtration). Here $\theta \in \mathbb{R}$ is a constant.

b) $\text{Cov}(B_s, B_t) = s \wedge t$

Proof Exercise.

$\dashv$

1.3 Characteristic Functions and Independence

Suppose that $(X_1, \ldots, X_n)$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that the characteristic function $\varphi_{X_1,\ldots,X_n} : \mathbb{R}^n \to \mathbb{C}$ is defined by

$$\varphi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \mathbb{E}[e^{i(t_1X_1+\cdots+t_nX_n)}] = \mathbb{E}[e^{i\mathbf{t}\cdot\mathbf{X}}]$$

The most important fact about characteristic functions is the following:

Theorem 1.3.1 (Lévy) *Two random vectors have the same distribution if and only if they have the same characteristic function.*

$\square$

The proof of the above theorem may be found in almost any advanced text on probability theory.

As a simple corollary, we have the following useful result:

Corollary 1.3.2 *Two random variables X, Y are independent if and only if $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$.*

Proof: If X, Y are independent, then

$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{i(sX+tY)} = \mathbb{E}[e^{isX}e^{itY}] = \mathbb{E}[e^{isX}] \cdot \mathbb{E}[e^{itY}] = \varphi_X(s)\varphi_Y(t)$$

Conversely, if suppose that $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$. Let $\bar{X}, \bar{Y}$ be independent random variables such that $\bar{X}$ has the same distribution as X, and $\bar{Y}$ as Y. Then $\varphi_{\bar{X}} = \varphi_X$ and $\varphi_{\bar{Y}} = \varphi_Y$. Thus

$$\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t) = \varphi_{\bar{X}}(s)\varphi_{\bar{Y}}(t) = \varphi_{\bar{X},\bar{Y}}(s,t)$$

as $\bar{X}, \bar{Y}$ are independent. It follows that (X, Y) and $(\bar{X}, \bar{Y})$ have the same characteristic function, and thus the same distribution. In particular, X is independent of Y, because $\bar{X}$ is independent of $\bar{Y}$.

$\square$

Theorem 1.3.3 (Kač) *A random vector* $\mathbf{X}$ *is independent of a σ–algebra $\mathcal{G}$ if and only if*

$$\mathbb{E}[e^{i\mathbf{t}\cdot\mathbf{X}}|\mathcal{G}] = \mathbb{E}[^{i\mathbf{t}\cdot\mathbf{X}}] \qquad \textit{all } \mathbf{t}$$

Proof: $(\Rightarrow)$ is a basic property of conditional expectation.
$(\Leftarrow)$ We will prove it for the one–dimensional case. Let Y be any $\mathcal{G}$–measurable random variable. Then, using the properties of conditional expectation,

$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{i(sX+tY)}] = \mathbb{E}[\ \mathbb{E}[e^{i(sX+tY)}|\mathcal{G}]\ ] = \mathbb{E}[e^{itY}\mathbb{E}[e^{isX}|\mathcal{G}]\ ] = \mathbb{E}[e^{itY}]\mathbb{E}[e^{isX}] = \varphi_X(s)\varphi_Y(t)$$

Hence X is independent of every $\mathcal{G}$–measurable random variable, and thus independent of $\mathcal{G}$.

$\square$

1.4 The Multi–Normal Distribution

We begin by gathering some results about the (multivariate) normal distribution. Recall:

Definition 1.4.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable $X : \Omega \to \mathbb{R}$ is normal (or Gaussian) if X has a density function of the form

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

i.e. if for all $A \in \mathcal{B}(\mathbb{R})$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(t)\, dt$$

□

It is well–known that

$$\mathbb{E}X = \mu \qquad \mathrm{Var}(X) = \sigma^2$$

and we say that X is $N(\mu, \sigma^2)$. It is also well–known that the characteristic function of X is

$$\phi_X(t) = \mathbb{E}[e^{itX}] = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$$

Definition 1.4.2 A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T : \Omega \to \mathbb{R}^d$ is called *multivariate Gaussian* or *(multi)normal* if $\mathbf{X}$ has a density of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)^T}{2}\right)$$

where μ is a d–dimensional vector, and Σ a symmetric positive definite $d \times d$–matrix[1]. It is easy to check that in that case

$$\mu = \mathbb{E}\mathbf{X} = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_d \end{pmatrix}$$

and

$$\Sigma = \mathbb{E}[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu)] = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_d) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \mathrm{Cov}(X_d, X_2) & \ldots & \mathrm{Cov}(X_d, X_d) \end{pmatrix}$$

We say that $\mathbf{X}$ has mean μ and covariance matrix Σ.

□

[1] A square matrix A is non–negative definite if $\mathbf{x}^T A\mathbf{x} \geq 0$ for any $\mathbf{x}$. It is positive definite if it is non–negative semidefinite and $\mathbf{x}^T A\mathbf{x} = 0$ only if $\mathbf{x} = 0$. Note that if $\mathbf{X}$ is a random vector with covariance matrix Σ, then $\mathbf{a}^T\Sigma\mathbf{a} = \mathrm{Var}(\mathbf{a}\cdot\mathbf{X}) \geq 0$, so that Σ is positive definite. That Σ is symmetric is obvious.

Next, we examine the characteristic function of a multinormal random vector. If $\mathbf{X} = (X_1, \ldots, X_d)$ is a random vector, its characteristic function $\varphi_{\mathbf{X}} : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}e^{i\mathbf{t}\cdot\mathbf{X}} = \int e^{i\mathbf{t}\cdot\mathbf{X}}\, d\mathbb{P} = \int e^{i\mathbf{t}\cdot\mathbf{x}}\ \mathbb{P}_{\mathbf{X}}(d\mathbf{x})$$

where $\mathbb{P}_{\mathbf{X}}$ is the distribution of $\mathbf{X}$, i.e. a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying $\mathbb{P}_{\mathbf{X}}(A) = \mathbb{P}(\mathbf{X} \in A)$.

Proposition 1.4.3 *If $\mathbf{X}$ is multinormal with mean vector μ and covariance matrix Σ, then its characteristic function is given by:*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}\cdot\mu - \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}}$$

Proof: An exercise in integration.

$\dashv$

Remarks 1.4.4 Some sources define a random vector X to be normal if and only if there is a non–negative definite matrix Σ such that the characteristic function of X is given as above. This extends the concept of a multinormal vector slightly, as it does not require Σ to be invertible.

For example, if X_1 is $N(0,1)$, and $X_2 = 0$ is constant, then $\varphi_{\mathbf{X}}(t_1, t_2) = e^{-\frac{1}{2}t_1^2}$, and so $\Sigma = \left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$. Then Σ is not invertible, and thus $\mathbf{X}$ is not multinormal in the sense of Definition 1.4.1, but it is multinormal in the extended sense.

$\square$

Proposition 1.4.5 *Let $X_i : \Omega \to \mathbb{R}$ for $i = 1, \ldots, d$. Then $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is multinormal if and only if*

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_d X_d$$

is (univariate) normal for all $\lambda_1, \lambda_2, \ldots, \lambda_d \in \mathbb{R}$.

Proof: Recall *Lévy's Inversion Theorem*: If two random vectors have the same characteristic functions, then they have the same distributions. Let μ be the mean and Σ the covariance matrix of $\mathbf{X}$.

Now suppose that $\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_d X_d$ is multinormal for all $\lambda = (\lambda_i) \in \mathbb{R}^d$, and define $Z_\lambda = \lambda \cdot \mathbf{X}$. Clearly

$$\mathbb{E}[Z_\lambda] = \lambda \cdot \mu \qquad \mathrm{Var}(Z_\lambda) = \lambda^T\Sigma\lambda$$

Now because Z_λ is univariate normal, its characteristic function is given by

$$\varphi_{Z_\lambda}(t) = e^{i\lambda\cdot\mu t - \frac{1}{2}\lambda^T\Sigma\lambda t^2}$$

$\dashv$

In particular, substituting $t = 1$, we get

$$\mathbb{E}[e^{iZ_\lambda}] = \varphi_{Z_\lambda}(1) = e^{i\lambda\cdot\mu - \frac{1}{2}\lambda^T\Sigma\lambda}$$

Now note that $\varphi_{\mathbf{X}}(\lambda) = \mathbb{E}[e^{i\lambda\cdot\mathbf{X}}] = \mathbb{E}[e^{iZ_\lambda}]$. It follows that $\mathbf{X}$ has the characteristic function of a multinormal random variable with mean μ and covariance matrix Σ. By Lévy Inversion, this means that $\mathbf{X}$ *is* a multinormal random variable with mean μ and covariance matrix Σ.

Conversely, suppose that $\mathbf{X}$ is multinormal (with mean μ and covariance matrix Σ), and let $\lambda \in \mathbb{R}^d$. Then $\varphi_{Z_\lambda}(t) = \mathbb{E}[e^{i\lambda\cdot\mathbf{X}t}] = \varphi_{\mathbf{X}}(\lambda t) = e^{i\lambda\cdot\mu t - \frac{1}{2}\lambda^T\Sigma\lambda t^2}$. Thus Z_λ has the characteristic function of a (univariate) normal random variable, and is therefore normal.

$\dashv$

It is now easy to prove the following:

Corollary 1.4.6 *If $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T$ is multinormal, then each X_d is normal.*

$\dashv$

The converse of this is false, as you will see in the next exercise.

Recall that two random variables X, Y are called **uncorrelated** provided that

$$\mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y$$

In that case their correlation coefficient ρ_{XY} is zero[2]. It is well–known that independent random variables are uncorrelated, but the converse is not true: Uncorrelated random variables need not be independent. However

Proposition 1.4.7 *Suppose that $\mathbf{X} = (X_1, \ldots, X_d)^T$ is a multinormal vector. Then $X_1, \ldots, X_d$ are mutually independent if and only if they are uncorrelated, i.e. if and only if the covariance matrix is diagonal.*

$\dashv$

Before we prove this proposition, it will be handy to note:

Lemma 1.4.8 *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector. Then $X_1, \ldots, X_d$ are independent if and only if*

$$\varphi_{\mathbf{X}}(t_1, \ldots, t_d) = \varphi_{X_1}(t_1) \cdot \cdots \cdot \varphi_{X_d}(t_d)$$

i.e. iff the characteristic function of the random vector can be factorized as a product of characteristic functions of the individual components.

Proof: For simplicity of notation, we prove this for the case $d = 2$. Clearly, if X, Y are independent, then $\varphi_{X,Y}(s, t) = \varphi_X(s)\varphi_Y(t)$.

[2] $\rho_{XY} = \frac{\mathrm{Covar}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$, and $\mathrm{Covar}(X, Y) = \mathbb{E}XY - \mathbb{E}X \cdot \mathbb{E}Y$.

Now suppose that $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$ for all $s, t \in \mathbb{R}$. Then

$$\begin{aligned}\int e^{i(sx+ty)}\, \mathbb{P}_{X,Y}(dx,dy) &= \varphi_{X,Y}(s,t) \\ &= \varphi_X(s)\varphi_Y(t) \\ &= \left(\int e^{isx}\, \mathbb{P}_X(dx)\right)\left(\int e^{ity}\, \mathbb{P}_Y(dy)\right) \\ &= \int e^{i(sx+ty)}\, d\mathbb{P}_X \otimes \mathbb{P}_Y(dx,dy)\end{aligned}$$

(using Fubini's Theorem). Here $\mathbb{P}_{X,Y}, \mathbb{P}_X, \mathbb{P}_Y$ are the distributions of the associated random vectors/variables. It follows that $\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$, and thus that X, Y are independent.

$\dashv$

Proof of 1.4.7 If the covariance matrix Σ is diagonal, then the characteristic function of $\mathbf{X}$ is easily seen to be factorisable in the sense of the previous lemma.

$\dashv$

1.5 Gaussian Processes

Definition 1.5.1 A stochastic process $(X_t : t \geq 0)$ is said to be *Gaussian* if and only if for any $0 \leq t_1 < \cdots < t_d$ the random vector $(X_{t_1}, \ldots, X_{t_d})$ is multivariate Gaussian.

$\square$

Proposition 1.5.2 *An a.s. continuous stochastic process X_t (with $X_0 = 0$) is a Brownian motion if and only if it is a Gaussian process with $\mathbb{E}X_t = 0$ (for all t) and* $\mathrm{Cov}(X_s, X_t) = s \wedge t$.

Proof: It is an exercise that if B_t is a Brownian motion, then $\mathrm{Cov}(B_s, B_t) = s \wedge t$. To see that Brownian motion is a Gaussian process, consider $0 \leq t_0 < \cdots < t_n$. We must show that the random vector $(B_{t_1}, \ldots, B_{t_n})$ is multivariate Gaussian. For this, it suffices to show that $\lambda_0 B_{t_0} + \cdots + \lambda_d B_{t_d}$ is normal for any $\lambda_0, \ldots, \lambda_n \in \mathbb{R}$. But $\lambda_0 B_{t_0} + \cdots + \lambda_n B_{t_n}$ can be rewritten as $\alpha_1(B_{t_1} - B_{t_0}) + \ldots \alpha_n(B_{t_n} - B_{t_{n-1}})$, a sum of *independent* normal random variables. But sums of independent normal variables are normal. Hence Brownian motion is a Gaussian process.

Conversely, suppose that X_t is an a.s. continuous Gaussian process with $\mathbb{E}X_t = 0$ and $\mathrm{Cov}(X_s, X_t) = s \wedge t$. Note that if $s < t$, then $\mathrm{Var}(X_t - X_s) = t - 2s + s = t - s$, so that $X_t - X_s \sim N(0, t-s)$. It remains to show that X_t has independent increments. So let $0 \leq t_0 < \cdots < t_n$, and let $Y_k = X_{t_k} - X_{t_{k-1}}$. To show that the Y_k are independent, it suffices to show that the covariance matrix Σ of the multinormal vector $\mathbf{Y} = (Y_1, \ldots Y_n)^T$ is diagonal. Now if $i < j$, then

$$\begin{aligned}\Sigma_{ij} &= \mathbb{E}Y_i Y_j \\ &= \mathrm{Cov}(X_{t_i}, X_{t_i}) - \mathrm{Cov}(X_{t_{i-1}}, X_{t_j}) - \mathrm{Cov}(X_{t_i}, X_{t_{j-1}}) + \mathrm{Cov}(X_{t_{i-1}}, X_{t_{j-1}}) \\ &= t_i - t_{i-1} - t_i + t_{i-1} \\ &= 0\end{aligned}$$

Because Σ is symmetric, we also have $\Sigma_{ij} = 0$ for $i > j$, i.e. $\Sigma_{ij} = 0$ whenever $i \neq j$. Hence Σ is a diagonal matrix.

$\dashv$

1.6 Some Useful and Interesting Properties

For those brought up on a diet of calculus and smooth functions, Brownian motion has many weird and counterintuitive properties, some of which are described below. (Indeed, in the 19th century, some of these would have been regarded as a "proof" of the non–existence of Brownian motion.)

Proposition 1.6.1 *Suppose that B_t is a standard Brownian motion. Then so are*

(1) $\hat{B}_t = cB_{t/c^2}$ *for* $c \neq 0$ *(Scaling)*

(2) $\hat{B}_t = tB_{\frac{1}{t}}$ *for* $t > 0$ *(Time Inversion)*
$\hat{B}_0 = 0$

(3) $\hat{B}_t = (B_{t+a} - B_a : t \geq 0)$ *for any* $a \geq 0$ *(Time Homogeneity)*

Proof: We prove only (2), and leave (1), (3) as exercises. Let $\hat{B}_t = tB_{\frac{1}{t}}$ for $t > 0$, and put $\hat{B}_0 = 0$. It is clear that $\hat{B}^t$ is a Gaussian process (because B_t is), and that $\mathbb{E}\hat{B}_t = 0$ for all t. Moreover, if $0 \leq s < t$, then $\mathrm{Cov}(\hat{B}_s, \hat{B}_t) = ts(\frac{1}{s} \wedge \frac{1}{t}) = \frac{ts}{t} = s = s \wedge t$. Thus the only thing that still needs to be proved is that $\hat{B}_t$ is continuous. This is certainly obvious for $t > 0$, because B_t is continuous. We need only prove, therefore, that $\hat{B}_t$ is continuous at $t = 0$, i.e. that $\lim_{t \downarrow 0} tB_{\frac{1}{t}} = 0$ a.s.

A quick way of seeing this is by the Strong Law of Large numbers: If N is large, then

$$B_N = (B_1 - B_0) + (B_2 - B_1) + \ldots (B_N - B_{N-1})$$

is the sum of a large number of identically distributed random variables. Thus $B_N/N \to 0$ a.s. as $N \to \infty$. Put $t = \frac{1}{N}$ and use continuity of $\hat{B}_u$ for $u > 0$ to conclude $t\hat{B}_{1/t} \to 0$ as $t \downarrow 0$.

A more roundabout way of seeing this is as follows: $\hat{B}_t \to 0$ as $t \downarrow 0$ if and only if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that $\sup_{0<s<\frac{1}{m}} |\hat{B}_s| < \frac{1}{n}$, and, by continuity, this is the case if and only if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that $\sup_{0<q<\frac{1}{m}, q \in \mathbb{Q}} |\hat{B}_q| < \frac{1}{n}$. Thus

$$\mathbb{P}(\lim_{t \downarrow 0} \hat{B}_t = 0) = \mathbb{P}(\bigcap_n \bigcup_m \bigcap_{0<q<\frac{1}{m}} \{|\hat{B}_q| < \frac{1}{n}\})$$

Now for some subtle reasoning: If we can show that $B_q, \hat{B}_q$ are identically distributed for $q > 0$, then

$$\mathbb{P}(\bigcap_n \bigcup_m \bigcap_{0<q<\frac{1}{m}} \{|\hat{B}_q| < \frac{1}{n}\}) = \mathbb{P}(\bigcap_n \bigcup_m \bigcap_{0<q<\frac{1}{m}} \{|B_q| < \frac{1}{n}\})$$

But the right–hand side is just $\mathbb{P}(\lim_{t \downarrow 0} B_t = 0)$, and this equals 1, since B_t is a.s. continuous at $t = 0$.

Hence if $B_q, \hat{B}_q$ are identically distributed for $q > 0$, then $\mathbb{P}(\lim_{t\downarrow 0} \hat{B}_t = 0) = \mathbb{P}(\lim_{t\downarrow 0} B_t = 0) = 1$, as required. But we know that $\hat{B}_t$ is Gaussian with the same means and covariances as B_t, so that $B_q, \hat{B}_q$ do indeed have the same finite–dimensional distributions. This concludes the proof.

$\dashv$

Note that the scaling property can also be usefully phrased as follows:

For any $a > 0$, $\frac{1}{\sqrt{a}} B_{at}$ is a Brownian motion.

You also have to be careful about Brownian motions relative to a filtration here. For example, if B_t is an $\mathcal{F}_t$–Brownian motion and $c > 1$, then cB_{t/c^2} is not a $\mathcal{F}_t$– Brownian motion: If $c > 1$, then $t/c^2 < t$, and thus $B_{t/c^2} - B_{s/c^2}$ need not be independent of $\mathcal{F}_s$. (Take $c = 2, t = 2, s = 1$ for example.)

Proposition 1.6.2 *A standard Brownian motion B_t has the following property:*

$$\mathbb{P}(\sup_{t\geq 0} B_t = +\infty, \inf_{t\geq 0} B_t = -\infty) = 1$$

Thus, with probability 1, a Brownian sample path will eventually exceed all bounds, positive and negative.

Proof: Let $Z = \sup_{t\geq 0} B_t$, and let $c > 0$. Recall that $\hat{B}_t := cB_{t/c^2}$ is a Brownian motion also — the scaling property. Similarly $\tilde{B}_t := B_{t+1} - B_1$ is a Brownian motion also — time homogeneity. Hence $\hat{Z} := \sup_t \hat{B}_t$, $\tilde{Z} := \sup_t \tilde{B}_t$, and Z all have the same distribution.

Now

$$\begin{aligned}
\mathbb{P}(Z \leq a) &= \mathbb{P}(\sup_t B_t \leq a) \\
&= \mathbb{P}(\sup_t B_{t/c^2} \leq a) \\
&= \mathbb{P}(\sup_t \hat{B}_t \leq ca) \\
&= \mathbb{P}(\hat{Z} \leq ca) \\
&= \mathbb{P}(Z \leq ca)
\end{aligned}$$

So for any $c > 0$ and any $a \in \mathbb{R}$, we have $\mathbb{P}(Z \leq a) = \mathbb{P}(Z \leq ca)$. It follows that $\mathbb{P}(Z \leq a) = \mathbb{P}(Z \leq b)$ for any $0 < a, b < \infty$, and thus $\mathbb{P}(0 < Z < \infty) = 0$. Since Z is necessarily non-negative (because $B_0 = 0$) we have $\mathbb{P}(Z = 0) + \mathbb{P}(Z = \infty) = 1$.

But $\mathbb{P}(Z = 0) \leq \mathbb{P}(B_1 \leq 0$ and $B_u \leq 0$ for all $u > 1)$ as $\{\sup_t B_t = 0\} \subseteq \{B_1 \leq 0\} \cap \{B_u \leq 0, \text{all } u > 1\}$. Now $\tilde{Z}$ has the same distribution as Z, and therefore takes values 0 and ∞ only. But if $\tilde{Z}(\omega) = \infty$, then obviously $\sup_{u\geq 1} B_u(\omega) = \infty$ also. Hence if $B_u(\omega) \leq 0$ for all $u \geq 1$, then $\tilde{Z}(\omega) = 0$, and hence $\{B_1 \leq 0, B_u \leq 0, \text{all } u > 1\} \subseteq \{B_1 \leq 0, \tilde{Z} = 0\}$. But as B_1 and $B_{t+1} - B_t$ are independent for all t, we see that $B_1, \tilde{Z}$ are independent. Hence $\mathbb{P}(Z = 0) \leq \mathbb{P}(B_1 \leq 0)\mathbb{P}(\tilde{Z} = 0) = \frac{1}{2}\mathbb{P}(Z = 0)$. We conclude that $\mathbb{P}(Z = 0) = 0$, so that $\mathbb{P}(Z = \infty = 1)$.

Similarly $\mathbb{P}(\inf_t B_t = -\infty) = 1$, and thus $\mathbb{P}(\sup_{t\geq 0} B_t = +\infty, \inf_{t\geq 0} B_t = -\infty) = 1$ as well.

$\dashv$

Corollary 1.6.3 *If $h > 0$, then $\mathbb{P}(B_t = 0 \text{ i.o. for } 0 \leq t \leq h) = 1$.*
Thus if a Brownian motion crosses the t–axis, it will do so again infinitely often in any succeeding time interval, no matter how small, with probability 1.

Proof: Exercise.
[Hint: Use the previous proposition, time inversion and time homogeneity.]

$\dashv$

Thus Brownian motion is extremely "bouncy", and this bouncyness is what leads to the difficulties in the definition of the stochastic integral later on.

Definition 1.6.4 A stochastic process X_t is said to be α*–self-similar* for some $\alpha > 0$ if and only if for any $c > 0$ and any $0 \leq t_1 < \cdots < t_n$ the rasnom vectors

$$c^\alpha (X_{t_1}, \ldots, X_{t_n}) \quad \text{and} \quad (X_{ct_1}, \ldots, X_{ct_n})$$

are identically distributed.

$\square$

Note that Brownian motion is $\frac{1}{2}$–self-similar, by the scaling property.

Proposition 1.6.5 *If a stochastic process X_t is α–self-similar and has stationary increments for some $0 < \alpha < 1$, then for any $t_0 > 0$ we have*

$$\limsup_{t \downarrow t_0} \frac{|X_t - X_{t_0}|}{t - t_0} = \infty$$

with probability 1.
It follows that X_t is not differentiable at t_0 with probability 1.

Proof: Because X_t is stationary, we may, by shifting if necessary, assume that $t_0 = 0$. Choose a countable sequence $t_n \downarrow 0$. By self-similarity, $0^\alpha X_t$ and X_{0t} are identically distributed. Thus $X_0 = 0$ a.s.

Now if $x > 0$, then

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty} \sup_{0 \leq s \leq t_n} \left| \frac{X_s}{s} \right| > x \right) &= \lim_{n\to\infty} \mathbb{P}\left(\sup_{0 \leq s \leq t_n} \left| \frac{X_s}{s} \right| > x \right) \\
&\geq \limsup_{n\to\infty} \mathbb{P}\left(\left| \frac{X_{t_n}}{t_n} \right| > x \right) \\
&= \limsup_{n\to\infty} \mathbb{P}\left(t_n^{\alpha-1} |X_1| > x \right) \\
&= 1
\end{aligned}$$

since $\alpha - 1 < 0$, so that $t_n^{\alpha-1} \uparrow \infty$. Thus $\mathbb{P}\left(\limsup_{n\to\infty} \left| \frac{X_{t_n}}{t_n} \right| = \infty \right) = 1$ for any sequence $t_n \downarrow 0$.

$\dashv$

Corollary 1.6.6 *Given $t_0 \geq 0$, then with probability 1, Brownian motion is not differentiable at t_0.*

$\dashv$

In fact, a stronger result is true:

Theorem 1.6.7 *Almost surely, a Brownian motion sample path is nowhere differentiable.*

This result is harder to prove, and can be omitted.

Proof: By the self–similarity of Brownian motion, it clearly suffices to prove the result on $[0, 1]$.

Suppose that $\omega \in \Omega$ and $t \geq 0$ are such that $B'_t(\omega)$ exists. Then both the left– and right derivatives exist (and are equal). We prove that the set of ω for which the right derivative exists at some t has measure 0.

Now if the right derivative of $B(\omega)$ exists at time t, then there is $N \in \mathbb{N}$ such that

$$\lim_{h \downarrow 0} \frac{|B_{t+h}(\omega) - B_t(\omega)|}{h} < N$$

i.e. $|B_{t+h}(\omega) - B_t(\omega)| \leq Nh$ for all sufficiently small $h > 0$. In particular, there exists an $m \in \mathbb{N}$ such that $0 \leq h \leq \frac{1}{m}$ implies $|B_{t+h}(\omega) - B_t(\omega)| \leq Nh$. For reasons that will become apparent in a moment, put $n = 4m$. Then we have shown that for any ω and any t, if the right derivative of $B(\omega)$ at t exists, then we can find N and n such that

$$0 \leq h \leq \frac{4}{n} \quad \text{implies} \quad |B_{t+h}(\omega) - B_t(\omega)| \leq Nh$$

Now define

$$A_{n,N} = \{\omega : |B_{t+h}(\omega) - B_t(\omega)| < Nh \text{ whenever} 0 \leq h \leq \frac{4}{n}\}$$

It is now clear that

$$\{\omega : \exists t (B'_t(\omega) \text{ exists})\} \subseteq \bigcup_N \bigcup_n A_{n,N}$$

Indeed, the set of ω for which $B(\omega)$ has a right derivative at some t is contained in the union on the right–hand side.

For future reference, note that $A_{n,N}$ is increasing in n, i.e.

$$A_{1,N} \subseteq A_{2,N} \subseteq \cdots \subseteq A_{n,N} \subseteq \ldots$$

To prove the theorem, it suffices to show that each $A_{n,N}$ has measure 0. So fix $n, N \in \mathbb{N}$ and define

$$\Delta_n(k) = B_{\frac{k+1}{n}} - B_{\frac{k}{n}}$$

Further define

$$k_t^n = \inf\{k \in \mathbb{N} : \frac{k}{n} \geq t\}$$

Note that $\frac{k_t^n - 1}{n} < t \leq \frac{k_t^n}{n} < \frac{k_t^n + 1}{n} < \frac{k_t^n + 2}{n} < \frac{k_t^n + 3}{n}$ and that $\frac{k_t^n + j}{n} - t < \frac{j+1}{n}$. By the triangle inequality, therefore,

$$|\Delta_n(k_t^n + j)| \leq \frac{7N}{n} \qquad \text{for } j = 0, 1, 2$$

For example,

$$\begin{aligned}|\Delta_n(k_t^n + 2)| &= |B_{\frac{k_t^n+3}{n}} - B_{\frac{k_t^n+2}{n}}| \\ &\leq |B_{\frac{k_t^n+3}{n}} - B_t| + |B_{\frac{k_t^n+2}{n}} - B_t| \\ &\leq N\left(\frac{4}{n} + \frac{3}{n}\right)\end{aligned}$$

It follows that

$$A_{n,N} \subseteq \bigcup_{k=0}^{n} \bigcap_{j=0}^{2} \left\{|\Delta_n(k+j)| \leq \frac{7N}{n}\right\}$$

and so

$$\mathbb{P}(A_{n,N}) \leq \sum_{k=0}^{n} \mathbb{P}\left(\bigcap_{j=0}^{2} \left\{|\Delta_n(k+j)| \leq \frac{7N}{n}\right\}\right)$$

But

$$\begin{aligned}&\mathbb{P}\left(\bigcap_{j=0}^{2} \left\{|\Delta_n(k+j)| \leq \frac{7N}{n}\right\}\right) \\ &= \mathbb{P}\left(|\Delta_n(k)| \leq \frac{7N}{n}\right)\mathbb{P}\left(|\Delta_n(k+1)| \leq \frac{7N}{n}\right)\mathbb{P}\left(|\Delta_n(k+2)| \leq \frac{7N}{n}\right) \\ &\qquad\qquad \text{(by independence of increments)} \\ &= \mathbb{P}\left(|Z| \leq \frac{7N}{\sqrt{n}}\right)^3 \quad \text{where } Z \sim N(0,1) \\ &= \left[\frac{2}{\sqrt{2\pi}}\int_0^{\frac{7N}{n}} e^{\frac{-x^2}{2}}\, dx\right]^3 \\ &\leq \left(\frac{14N}{\sqrt{2\pi n}}\right)^3 \qquad \text{because } e^{\frac{-x^2}{2}} \leq 1\end{aligned}$$

Hence

$$\mathbb{P}(A_n, N) \leq (n+1)\left(\frac{14N}{\sqrt{2\pi n}}\right)^3 \to 0 \quad \text{as } n \to \infty$$

But since $\mathbb{P}(A_{n,N}) \subseteq \mathbb{P}(A_{n+1,N} \subseteq \mathbb{P}(A_{n+2,N}) \subseteq \ldots$, it is clear that we must have $\mathbb{P}(A_{n,N}) = 0$ for all n, N.

$\dashv$

1.7 Exercises

Exercise 1.7.1 Write a program to draw a random walk $R_t := \sum_{k=1} nX_k$, where $t = n\Delta t$, and $\mathbb{P}(X_k = \pm\Delta x) = \frac{1}{2}$. Play around with the relationship between Δx and Δt and see what happens as $\Delta t \to 0$. (e.g. try $\Delta x = \sqrt{\Delta t}, \Delta t, \Delta t^2, \sqrt[3]{\Delta t}, 0.2\sqrt{\Delta t}$ etc.).

Exercise 1.7.2 Show that if B_t is a standard Brownian motion, then $\text{Cov}(B_s, B_t) = s \wedge t$ (where $s \wedge t := \min\{s, t\}$.)

Exercise 1.7.3 Suppose that B_t is a standard Brownian motion. Then so are

(1) $\hat{B}_t = cB_{t/c^2}$ for $c \neq 0$.

(2) $\hat{B}_t = B_{t+a} - B_a$ for any $a \geq 0$.

Exercise 1.7.4 Show that if $h > 0$, then $\mathbb{P}(B_t = 0 \ \text{ infinitely often for } 0 \leq t \leq h) = 1$.
[Hint: Use Proposition 1.6.2 and time inversion.]

Exercise 1.7.5 Suppose that B_t is a standard 1–dimensional Brownian motion. Prove that B_t, $B_t^2 - t$ and $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$ are all martingales (with respect to the natural filtration). Here $\theta \in \mathbb{R}$ is constant.

Exercise 1.7.6 Give an example of two uncorrelated normally distributed random variables X, Y that are *not independent*.
[Hint: Let $a > 0$ be an as yet unspecified real, and let $X \sim N(0, 1)$. Define Y as follows:

$$Y = \begin{cases} X & \text{if } |X| \leq a \\ -X & \text{else} \end{cases}$$

Show that $Y \sim N(0, 1)$ as well. Further,

$$\mathbb{E}XY = \frac{2}{\sqrt{2\pi}} \left[\int_0^a x^2 e^{-x^2/2}\, dx - \int_a^\infty x^2 e^{-x^2/2}\, dx \right]$$

Now show that it is possible to choose a so that $\mathbb{E}XY = 0$. Show that X, Y are uncorrelated random variables, but that X, Y are not independent. Conclude that (X, Y) is not bivariate normal.]

Exercises 1.7.7 (1) Show that if X, Y are independent normal random variables, then (X, Y) is bivariate normal. Conclude that $X + Y$ is normal. as well.

(2) Show that the a.s. limit of multivariate Gaussian vectors is multivariate Gaussian.

[Hint: (2) Suppose that $X^n = (X_1^n, \ldots, X_k^n)$ and that $X^n \to X$ a.s. You must show that $X = (X_1, \ldots, X_k)$ is multivariate Gaussian. Clearly $X^n \to X$ in distribution as well. Let φ_n be the characteristic function of X^n and let φ be the characteristic function of X. Because weak convergence is equivalent to pointwise convergence of characteristic functions, we have $\varphi_n(\theta) \to \varphi(\theta)$ for all $\theta \in \mathbb{R}^k$. From the structure of φ_n, conclude, using Lévy's Inversion Theorem, that X is also multivariate normal.]

Exercise 1.7.8 Calculate $\mathbb{E}[B_s|B_t = a]$ and $\text{Var}(B_s|B_t = a)$ in the case:

(a) $t \le s$

(b) $t \ge s$

Interpret your result in (b) geometrically.
[Hint: (b) If $f_{X,Y}(x,y)$ is the joint density function of (X,Y), then

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

is the density function of X given $Y=y$.]

Exercise 1.7.9 Show that the a.s. limit of multivariate Gaussian vectors is multivariate Gaussian.
What about convergence in probability and weak convergence?
[Hint: Suppose that $X^n = (X_1^n, \dots, X_k^n)$ and that $X^n \to X$ a.s. You must show that $X = (X_1, \dots, X_k)$ is multivariate Gaussian. Clearly $X^n \to X$ in distribution as well. Let φ_n be the characteristic function of X^n and let φ be the characteristic function of X. Because weak convergence is equivalent to pointwise convergence of characteristic functions, we have $\varphi_n(\theta) \to \varphi(\theta)$ for all $\theta \in \mathbb{R}^k$. From the structure of φ_n, conclude, using Lévy Inversion, that X is also multivariate normal.]

Exercise 1.7.10 On pg. 10 - 12 of *Diffusions, Markov Processes and Martingales*, by Rogers & Williams there is a construction of Brownian motion, due to Ciesielski. The aim of this exercise is to work through that proof, which is reproduced below. Ciesielski's method constructs a Brownian motion on the unit interval $[0,1]$. The shifting property of Brownian motion then allows it to be extended to all of $\mathbb{R}^+$.

Take some probability space on which there is defined an infinite sequence of independent $N(0,1)$ random variables. For reasons that will soon be apparent, we assume that they are indexed as $\{Z_{k,n} : n \in \mathbb{Z}^+, k \text{ odd}, k \le 2^n\}$ Now define

$$g_{1,0}(t) = 1$$

$$g_{k,n}(t) = \begin{cases} 2^{(n-1)/2} & \text{if } (k-1)2^{-n} < t \le k2^{-n} \\ -2^{(n-1)/2} & \text{if } k2^{-n} < t \le (k+1)2^{-n} \\ 0 & \text{else} \end{cases}$$

for $n \ge 1, k \le 2^n$, k odd. For notational convenience, let $S_n = \{(k,n) : k \text{ odd}, k \le 2^n\}$, $S = \bigcup_{n \ge 0} S_n$. The first thing to notice is that $\{g_{k,n} : (k,n) \in S\}$ *is a complete orthonormal system in* $L^2[0,1]$. The orthonormality of $g_{k,n}$ is easy to check; and, for completeness, if $f \in L^2[0,1]$ were orthogonal to all the $g_{k,n}$, then $F(t) = \int_0^t f(u)\, du$ would vanish at 0 and 1 (since $f \perp g_{1,0}$; and also at $\frac{1}{2}$ (since $f \perp g_{1,1}$); and also at $\frac{1}{4}$, $\frac{3}{4}$, (since $f \perp g_{1,2}, g_{3,2}$), ... Thus $F = 0$, and $f = 0$.

Now define $f_{k,n}(t) = \int_0^t g_{k,n}(u)\, du$, and the approximations $B_n(\cdot)$ to Brownian motion by

$$B_n(t) = \sum_{m=0}^{n} \sum_{(k,m)\in S_m} Z_{k,m} f_{k,m}(t)$$

Let us describe what these approximations are doing.
The first approximation B_0 is simply $tZ_{1,0}$, a straight line. The next approximation is obtained by adding on a Gaussian multiple of $f_{1,1}$, which is a tent–shaped function, vanishing at 0 and 1. The next approximation is obtained by adding on two Gaussian multiples of tent–shaped functions, which both vanish at $0, \frac{1}{2}$ and $1\ldots$.
The next stage of the proof is to establish that *the B_n converge uniformly* almost surely. Indeed, for any positive constant a_n,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{0\le t\le 1} |B_n(t) - B_{n-1}(t)| > a_n\Big)\\
&= \mathbb{P}\Big(\sup_{k} |Z_{k,n}| > 2^{(n+1)/2} a_n\Big)\\
&\qquad\qquad\text{(since the } f_{k,n} \text{ are all at most } 2^{-(n+1)/2})\\
&\le 2^{n-1}\mathbb{P}(|Z_{1,n}| > 2^{(n+1)/2} a_n)\\
&\le (4\pi)^{-1/2} 2^{n/2} a_n^{-1} \exp(-a_n^2 2^n)
\end{aligned}$$

by the elementary estimate

$$\int_x^\infty \exp(-\frac{1}{2}y^2)\, dy \le x^{-1} \exp(-\frac{1}{2}x^2)$$

We now aim to choose the constants a_n in such a way that

$$\sum_n 2^{n/2} a_n^{-1} \exp(-a_n^2 2^n) < \infty$$

$$\sum_n a_n < \infty$$

The first of these conditions will ensure that, almost surely,

$$\sup_{0\le t\le 1} |B_n(t) - B_{n-1}(t)| \le a_n \qquad \text{for all large enough } n;$$

the second will guarantee that the B_n converge uniformly (almost surely) to a limit B, which is therefore continuous. But these conditions are satisfied by the choice $a_n = (n2^{-n})^{1/2}$, for example.
Thus we have proved that, almost surely, the B_n converge to some continuous limit B, which we now must show is Brownian motion. As we saw…, the simplest way to do this is to check that B is a zero–mean Gaussian process with covariance structure $\mathbb{E}(B_s B_t) = s\wedge t$.

Obviously, each B_n is a zero–mean Gaussian process: the vector $(B_n(t_1), \ldots, B_n(t_k))$ is multivariate Gaussian. This converges almost surely (and so in distribution) to $(B(t_1), \ldots, B(t_k))$, and the limit of the covariances of the B_n gives the covariance of B. But

$$\mathbb{E}[B_n(s)B_n(t)] = \sum_{m=0}^{n} \sum_{(k,m)\in S_m} f_{k,m}(s) f_{k,m}(t)$$

by the independence of the $Z_{k,m}$, and this converges as $n \uparrow \infty$ to

$$\begin{aligned}
&\sum_{(k,m)\in S} f_{k,m}(s) f_{k,m}(t) \\
&= \int_0^1 I_{[0,s]}(u) I_{[0,t]}(u)\, du \\
&= s \wedge t
\end{aligned}$$

since $f_{k,m}(s) = \int_0^1 I_{[0,s]}(u) g_{k,m}(u)\, du$ is the Fourier coefficient of $g_{k,m}$ in the representation of $I_{[0,s]}$ in terms of the complete orthonormal system $\{g_{k,n} : (k,n) \in S\}$. Parseval's identity concludes the proof.

(a) Draw graphs of the functions $g_{1,0},\ g_{1,1},\ g_{1,2},\ g_{3,2},\ g_{1,3}, \ldots, g_{7,3}$ to get a feeling of what these functions look like.

(b) It is asserted in the proof that the family $\{g_{k,n} : (k,n) \in S\}$ is a complete orthonormal system in the Hilbert space $\mathcal{L}^2[0,1] = \mathcal{L}^2([0,1], \mathcal{B}[0,1], \lambda)$, where $\mathcal{B}[0,1]$ is the family of Borel subsets of the compact interval $[0,1]$ and λ is Lebesgue measure.

(i) Show that the $g_{k,n}$ are orthonormal, i.e. show that

$$\langle g_{k,n}, g_{l,m} \rangle = \begin{cases} 0 \text{ if } (k,n) \neq (l,m) \\ 1 \text{ else} \end{cases}$$

Here $\langle f, g \rangle$ is the usual inner product in $\mathcal{L}^2[0,1]$, i.e.

$$\langle f, g \rangle = \int_0^1 fg\, d\lambda$$

(ii) In a general Hilbert space, an orthonormal system $\{u_n : n \in \mathbb{N}\}$ is called complete if and only if for every u, we have

$$u = \sum_{n=0}^{\infty} \langle u_n, u \rangle u_n$$

i.e. the above series converges, and converges to u. Note that $\langle u_n, u \rangle$ is just the projection of u onto the unit vector u_n. Show that a system of orthonormal vectors is complete if and only if

$$\langle u_n, u \rangle = 0 \text{ for all } \iff u = 0$$

(iii) Haul out some old notes on Fourier series and identify the complete orthonormal system of functions used there. Write the Fourier coefficients of a continuous function f on a compact interval as inner products. It is for this reason that the projections $\langle u_n, u\rangle$ are called the *generalized Fourier coefficients* of u (w.r.t the basis $\{u_n\}$).

(iv) Now show that the set of $g_{k,n}$ is, in fact, a complete orthonormal system.

(c) Draw the graphs of the first few tent-shaped functions $f_{1,0}, \ldots, f_{7,3}$.

(d) In the proof that the approximations B_n converge uniformly almost surely, the following inequality is used:

$$\int_x^\infty \exp(-y^2/2)\, dy \le x^{-1} \exp(-x^2/2)$$

Prove that this inequality holds.

(e) Verify that the choice of the constants $a_n = (n2^{-n})^{1/2}$ will satisfy the two conditions asserted in the proof.

(f) Now that we know that the B_n converge uniformly (almost surely) to something, it remains to prove that that *something* is a Brownian motion, i.e. a continuous Gaussian process B with $\mathrm{Cov}(B_s, B_t) = s \wedge t$.

(i) Why is the limit B of the B_n continuous a.s.?

(ii) "Obviously, each B_n is a zero–mean Gaussian process." Why?

(iii) Show that

$$\mathrm{Cov}(B(s), B(t)) = \lim_{n\to\infty} \mathrm{Cov}(B_n(s), B_n(t)) = \sum_{(k,m)\in S} f_{k,m}(s) f_{k,m}(t)$$

(iv) The indicator functions $I_{[0,t]}$ can be represented as a generalized Fourier series

$$I_{[0,t]}(u) = \sum_{(k,n)\in S} b_{k,n} g_{k,n}(u)$$

because the system $\{g_{k,n}\}$ is complete. Show that the generalized Fourier coefficients of $I_{[0,t]}$ are given by

$$b_{k,m} = f_{k,m}(t)$$

(v) Now note that

$$s \wedge t = \int_0^1 I_{[0,t]}(u) I_{[0,s]}(u)\, du$$

and prove that

$$\sum_{(k,m)\in S} f_{k,m}(t) f_{k,m}(s) = s \wedge t$$

(vi) We now know that B is a continuous process with the right covariance structure. It remains to show that B is, in fact, a Gaussian process. Use Exercise 1.7.9 to accomplish this.

The proof is now complete.

Chapter 2

Martingale Theory

2.1 Elements of Discrete–Time Martingales

Martingales are amongst the most important objects in probability theory, and an entire sub–discipline of finance is based on them. Brownian motion is the most important continuous–parameter martingale, and is heavily used in financial modelling. In this chapter we first introduce the basic results about discrete–time martingales at a leisurely rate, taking time to build up intuition and facility with martingale calculations. In the next chapter, we will tackle continuous–parameter martingales.

2.1.1 Stochastic Processes and Filtrations

Informally, a (discrete–parameter) stochastic process X is a family of random variables indexed by a discrete time set, i.e. $X = X_1, X_2, X_3, \ldots$. The idea is that these model the outcomes of a series of random phenomena, such as the closing values of the S&P500. The X_n are thus successive values of some quantity under consideration. Note that the times of the random variables may not be evenly distributed in physical time; for example, the share index is recorded only on trading days.

We assume that the stochastic process $X = (X_n : n \in I)$ is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The time index set I will usually be the set of natural numbers, or the set of non–negative reals, or some finite initial segment these. For a particular outcome $\omega \in \Omega$, the sequence $X_1(\omega), X_2(\omega), \ldots$ is called a *sample path* of the process. Note that *one* outcome/state–of–the–world ω determines the values of all the X_n. We only know the value of X_n at time n, and so as time n increases, so does our knowledge of the state of the world. Since information is organised in σ–algebras, we associate with each time n a σ–algebra $\mathcal{F}_n$ modelling the knowledge at time n. We also assume that no information is lost or forgotten, so that information available at time n is also available at a later time $m > n$. This simply means that $\mathcal{F}_m \supseteq \mathcal{F}_n$. We thus model the flow of information as follows:

Definition 2.1.1 Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. An increasing sequence

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots \subseteq \mathcal{F}$$

of σ–algebras on Ω is called a **filtration**. We shall always assume that $\mathcal{F}_0$ contains all the sets of measure 0.
We also define

$$\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n) \subseteq \mathcal{F}$$

$\mathcal{F}_n$ represents the available information at time n, i.e. it contains all events A for which it is possible to decide at time n whether A has occurred or not.

Suppose that S_t is the share price at time t. We know S_2 a**t time $t = 2$. Thus each of the following events can be decided at time $t = 2$: Whether or not $X_2 = 5.00$; whether or not X_2 lies between 13.50 and 15.76, etc. It therefore follows that X_2 must be $\mathcal{F}_2$–measurable, i.e. that $\sigma(X_2) \subseteq \mathcal{F}_2$. Moreover, X_1 is also known at $t = 2$, so $\sigma(X_1, X_2) \subseteq \mathcal{F}_2$. However, at $t = 2$ we do not know the share price at time $t = 3$. Thus X_3 is not $\mathcal{F}_2$–measurable, although it is, of course $\mathcal{F}_3$–measurable.

In essence, to model the fact that the value of X_m is known at a later time n, we need to add the restriction that X_m is $\mathcal{F}_n$–measurable for all $n \geq m$. This just means that $\sigma(X_1, \ldots, X_n) \subseteq \mathcal{F}_n$, and so we define:

Definition 2.1.2 A stochastic process $X = (X_n, n \in I)$ is said to be *adapted* to a filtration $\mathcal{F}_n, n \in I$ provided that each X_n is $\mathcal{F}_n$–measurable. It follows trivially that this is the case if and only if

$$\sigma(X_1, \ldots, X_n) \subseteq \mathcal{F}_n$$

Exercise 2.1.3 Make sure that you can prove this trivial result.

$\square$

Note that to say that X is adapted to $\mathcal{F}_n$ simply means that the random variables X_n do not contain more information than the $\mathcal{F}_n$, although they may contain strictly less.

Note also that $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is the smallest filtration with respect to which X is adapted, i.e. that if X is also adapted to a filtration $\mathcal{G}_n$, then $\mathcal{F}_n \subseteq \mathcal{G}_n$. The filtration $\mathcal{F}_n$ contains just the information in the values of X up to time n, and is called the *natural* or *canonical* filtration of X. It contains just as much information as is contained in the X_n, and no more.

2.1.2 Martingales, Submartingales, Supermartingales

Martingales model a fair game, submartingales a favourable game, and supermartingales an unfavourable game. Here is the definition:

Definition 2.1.4 A stochastic process $X = (X_n : n \in \mathbb{N})$ is called a **supermartingale** (respectively **submartingale**) with respect to a filtration $\mathcal{F}_n, n \in \mathbb{N}$ if and only if

(a) Each $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$[a]

(b) X is adapted to $\mathcal{F}_n, n \in \mathbb{N}$.

(c) $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$ (respectively, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$) for each $n \in \mathbb{N}$

A **martingale** is simultaneously a sub– and a supermartingale, i.e. it satisfies $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for each $n \in \mathbb{N}$.
When we say that X is a (super/sub–)martingale, but we don't mention a specific filtration, then the natural filtration should be used.

[a] i.e. each X_n is *integrable*, which just means that $\mathbb{E}X_n$ exists, and is finite.

(Note that we've taken $\mathbb{N}$ as the index set. You shouldn't have any trouble generalizing the definition to the case where the index set is some initial segment $\{0, 1, 2, \ldots, T\}$ of $\mathbb{N}$).

Think of X_n as your total fortune after the n^{th} round of a gambling game. If X is a supermartingale, your expected fortune at time $n+1$ is less than your fortune at time n. It follows that this particular game is unfavourable, i.e. that you are likely to lose. If X is a martingale, then your expected fortune equals your present fortune: You are just as likely to win as to lose, and the game is fair.

Examples 2.1.5 (a) Suppose that the $X_n, n \in \mathbb{N}$ are independent random variables with $\mathbb{E}X_n = 0$, and that $\mathcal{F}_n, n \in \mathbb{N}$ is the natural filtration. Define $S_n = X_1 + \cdots + X_n$. Clearly $S = (S_n : n \in \mathbb{N})$ is a stochastic process adapted to $\mathcal{F}_n, n \in I$, and each S_n is integrable. Moreover,

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_1|\mathcal{F}_n] + \cdots + \mathbb{E}[X_n|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n]$$

Since X_m is $\mathcal{F}_n$–measurable for $m \leq n$, it follows that $\mathbb{E}[X_m|\mathcal{F}_n] = X_m$ if $m \leq n$. Moreover, since the X_m are independent random variables, X_{n+1} is independent of $\mathcal{F}_n$, and thus we have $\mathbb{E}[X_{n+1}|\mathcal{F}_m] = \mathbb{E}X_{n+1} = 0$. Hence

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = X_1 + \cdots + X_n + 0 = S_n$$

which proves that $S_n, n \in \mathbb{N}$ is a martingale.

(b) If we have the same situation as in (a), but with $\mathbb{E}X_n > 0$ for all n, then $S_n, n \in \mathbb{N}$ is a submartingale.

(c) If $X_n, n \in \mathbb{N}$ are random variables with the same mean $\mu = 0$ and the same variance σ^2, and if $S_n = X_1 + X_2 + \cdots + X_n$, then the process $W_n = S_n^2 - n\sigma^2$ is a martingale with respect to the natural filtration of the X_n. First note that each W_n is integrable if and only if S_n^2 is, but this follows because the variances $\sigma^2 = \mathbb{E}X_n^2$ exist, so that each $X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. To verify the martingale property, observe that $S_{n+1}^2 = S_n^2 + 2S_nX_{n+1} + X_{n+1}^2$. Further observe that $\mathbb{E}[S_nX_{n+1}|\mathcal{F}_n] = S_n\mathbb{E}[X_{n+1}|\mathcal{F}_n]$, because S_n is $\mathcal{F}_n$–measurable, and that

$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}X_{n+1} = 0$, since X_{n+1} is independent of $\mathcal{F}_n$. Thus:

$$\begin{aligned}\mathbb{E}[W_{n+1} - W_n|\mathcal{F}_n] &= \mathbb{E}[(S_n + X_{n+1})^2 - S_n^2 - \sigma^2)|\mathcal{F}_n] \\ &= 2\mathbb{E}[S_n X_{n+1}|\mathcal{F}_n] + \mathbb{E}[X_{n+1}^2|\mathcal{F}_n] - \sigma^2 \\ &= 2S_n\mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}X_{n+1}^2 - \sigma^2 \\ &= 2S_n \cdot \mathbb{E}X_{n+1} + \mathrm{Var}(X_n) - \sigma^2 \\ &= 0\end{aligned}$$

(d) Suppose that X_n are non-negative random variables with $\mathbb{E}X_n = 1$. Put $M_0 = 1$, and define

$$M_n = X_1 \cdot X_2 \cdot \dots \cdot X_n$$

for $n > 0$. Assume that each M_n is integrable. It is left as an exercise to show that M_n is a martingale.

(e) Consider a random walk. If it is symmetric, it is a martingale. If the probability p of going up is < 0.5, it is a supermartingale.

(f) One more interesting martingale demonstrates the accumulation of information about the value of a random variable over time. Let Y be an integrable random variable (i.e. $\mathcal{F}$–measurable). We do not necessarily know the value of Y at time n — there may not be enough information available. However, as time passes, we expect that our estimate will become more accurate. At time n, the best available approximation to Y is $Y_n = \mathbb{E}[Y|\mathcal{F}_n]$. We now show that Y_n is a martingale (with respect to the natural filtration). Firstly,

$$\mathbb{E}Y_n = \mathbb{E}[\mathbb{E}(Y|\mathcal{F}_n)] = \mathbb{E}Y$$

by the "Tower Property". This shows that each Y_n is integrable if Y is. Next,

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[Y|\mathcal{F}_n] = Y_n$$

by the Tower Property again. This proves the result.

What this means is that there are no trends in our estimates of Y. At each new time step, our revised estimate is just as likely to go up as it is to go down, and is expected to remain at the same value as our previous estimate. This makes sense: If we *expected* our estimates to increase, for example, then our estimates would not have been the best available. We ought to have built the expectation of increase into our estimates already.

(g) Note that if X_n is a martingale, and if φ is a *convex* function, then $\varphi(X_n)$ is a submartingale. Indeed,

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = \varphi(X_n)$$

by Jensen's inequality. It follows that if X_n is a martingale and if $p > 1$, then $|X_n|^p$ is a submartingale.

$\square$

Remarks 2.1.6 (a) If $X_n, n \in \mathbb{N}$ is a martingale, then $\mathbb{E}X_n = \mathbb{E}X_0$ for all n, i.e. all the X_n have the same mean. This is an easy exercise.

(b) We have defined the martingale property with respect to a filtration. Thus if X_n is a martingale with respect to one filtration, it may not be with respect to another. However, if X_n is a martingale with respect to some filtration $\mathcal{G}_n$, and if $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is the natural filtration, then X_n is also a martingale with respect to $\mathcal{F}_n$. To see this, first note that each X_n is $\mathcal{G}_n$–measurable (because X_n is adapted to $\mathcal{G}_n$ — part of the definition of *martingale*). Thus $\mathcal{F}_n \subseteq \mathcal{G}_n$ for each n. It now follows by the Tower Property that

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{G}_n]|\mathcal{F}_n] = \mathbb{E}[X_n|\mathcal{F}_n] = X_n$$

The last equality holds by "Taking out what is known", because X_n is $\mathcal{F}_n$–measurable. It is now not hard to see that if X_n is a martingale with respect to one filtration, it will also be a martingale with respect to any poorer (in information) filtration to which it is adapted.

(c) The converse of (b) is not true: If X_n is a martingale with respect to the natural filtration, it may not be a martingale with respect to a richer (in information) filtration. Find a simple example!

(d) Note that if X_n is a martingale, and if $m \geq n$, then $\mathbb{E}[X_m|\mathcal{F}_n] = X_n$. This is left as another exercise in the use of the Tower Property.

□

The following exercise will prove extremely useful:

Exercise 2.1.7 (Orthogonality of Martingale Increments)
Prove that if M_n is a martingale, then

$$\mathbb{E}[(M_n - M_m)^2|\mathcal{F}_k] = \mathbb{E}[M_n^2 - M_m^2|\mathcal{F}_k] \qquad k \leq m \leq n$$

Deduce that

$$\mathbb{E}[M_n]^2 = \mathbb{E}M_0^2 + \sum_{m=1}^{n} \mathbb{E}[(M_m - M_{m-1})^2]$$

□

2.1.3 Games and Strategies

Suppose that you take part in a game of chance, e.g. a game of coin tossing, roulette, or investing in the stock market. The game is repeated many times, and you place a bet each time. Let $\xi_n, n \in \mathbb{N}$ be a sequence of integrable random variables which represent your winnings (or losses, if negative) per unit stake in the n^{th} game. Thus, if you had wagered a stake C_n on the n^{th} game, you would have won $C_n \xi_n$.

If you played unit stakes all the way through, your total winnings after the n^{th} game would be

$$S_n = \xi_1 + \cdots + \xi_n \qquad \text{for } n \geq 1$$

Note that $S_0 = 0$, because you haven't won or lost anything yet.

If the game is *fair* then your chance of winning is the same as your chance of losing, and thus $\mathbb{E}\xi_n = 0$. In that case, S_n is a martingale with respect to the natural filtration

$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n) = \sigma(S_1, \dots, S_n)$. Similarly, if the game is unfavourable to you, then at time n you expect your winnings at time $n+1$ to be less than your current winnings, i.e. $\mathbb{E}[S_{n+1}|\mathcal{F}_n] \leq S_n$. Thus an unfavourable game is modelled by a supermartingale. A favourable game will clearly be modelled by a submartingale.

Suppose now that you have a *system*, i.e. a gambling strategy, which tells you when to bet, how much to bet etc. Your system, call it C, will tell you what stake C_n you should place on the n^{th} game. We allow negative stakes as well (which are essentially bets that you will lose)[1]. In that case, your total winnings after the n^{th} game will be

$$W_n = C_1\xi_1 + \cdots + C_n\xi_n$$

Now note that $\xi_n = S_n - S_{n-1} = \Delta S_n$, and thus that

$$W_n = \sum_{k=1}^{n} C_k(S_k - S_{k-1}) = \sum_{k=1}^{n} C_k \Delta S_k$$

which looks like a Riemann–Stieltjes sum. Your strategy $C = (C_n : n \in I)$ is also a stochastic process, but since we have to decide what stake to wager before the outcome of the n^{th} game is known, we must be able to decide the value of C_n on the basis of information available at time $n-1$ (i.e. after the $(n-1)^{\text{th}}$ game). Thus each C_n is $\mathcal{F}_{n-1}$–measurable. We have a name for this:

Definition 2.1.8 A stochastic process C is called *previsible* (or *non–anticipative*, or *predictable*), with respect to a filtration $\mathcal{F}_n$ provided that each C_n is $\mathcal{F}_{n-1}$–measurable, for $n \geq 1$. Note that C_0 is not defined.

Thus a gambling strategy is just a previsible process.

Consider an arbitrary adapted stochastic process Y_n. Then in general Y_n may exhibit both purely random behaviour and long–term trends. For example, for supermartingales the long–term trend is that it tends to decrease. Purely random behaviour is described by martingales, and trends are known beforehand, i.e. are previsible. We thus attempt to decompose Y_t into a martingale part and a previsible part, i.e. we try to write

$$Y_n = M_n + A_n$$

where M_n is a martingale, with $M_0 = Y_0$, and A_n is previsible, with $A_0 = 0$. In engineering, A_n is called the *signal*, and M_n the *noise*.

Suppose that we can actually find such a decomposition. We would then have

$$Y_{n+1} - Y_n = (M_{n+1} - M_n) + (A_{n+1} - A_n)$$

Taking conditional expectations immediately yields

$$A_{n+1} - A_n = \mathbb{E}[Y_{n+1}|\mathcal{F}_n] - Y_n$$

so that

$$M_{n+1} - M_n = Y_{n+1} - \mathbb{E}[Y_{n+1}|\mathcal{F}_n]$$

We now use this pair of equations to *define* M_n and A_n in the next theorem.

[1] We need negative stakes to model short sales, which are essentially just bets that a stock will lose value

Theorem 2.1.9 (Doob Decomposition Theorem)
Every process Y_n has a unique decomposition

$$Y_n = M_n + A_n$$

where M_n is a martingale with $M_0 = Y_0$, and A_n is previsible, null at $n = 0$. Moreover, if Y_n is a supermartingale, then A_n is decreasing.

Proof: Define M_n, A_n inductively by

$$\begin{cases} M_0 = Y_0, & M_{n+1} = M_n + Y_{n+1} - \mathbb{E}[Y_{n+1}|\mathcal{F}_n] \\ A_0 = 0, & A_{n+1} = A_n + \mathbb{E}[Y_{n+1}|\mathcal{F}_n] - Y_n \end{cases}$$

It is clear that M_n is a martingale and that A_n is previsible. Moreover

$$Y_m - Y_{m-1} = (M_m - M_{m-1}) + (A_m - A_{m-1})$$

summing over m from $m = 1$ to $m = n$ yields

$$Y_n = M_n + A_n$$

as required.

To see that this decomposition is unique, suppose that $Y_n = M_n' + A_n'$ is another decomposition with the same properties. We show by induction on n that $M = M', A = A'$: Note that $M_0 = M_0'$ by definition. Suppose that $M_n = M_n'$, and, consequently, that $A_n = A_n'$. Then

$$M_{n+1} - M_{n+1}' = A_{n+1} - A_{n+1}'$$

Taking conditional expectations with respect to $\mathcal{F}_n$, we obtain

$$0 = M_n - M_n' = A_{n+1}' - A_{n+1}$$

because A, A' are previsible. Hence $A_{n+1} = A_{n+1}'$, and so $M_{n+1} = M_{n+1}'$ as well. By induction, we have $M_n = M_n', A_n = A_n'$ for all $n \in \mathbb{N}$. This proves that the Doob Decomposition is unique.

If X_n is a supermartingale, then $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$, so the definition of A_{n+1} implies that $A_{n+1} \leq A_n$.

⊣

Exercise 2.1.10 Suppose that Y_t is a martingale. By Jensen's inequality, Y_t^2 will be a submartingale, and thus have an increasing trend. The previsible trend part A_t of Y_t^2 is called the *quadratic variation*, for the following reason:

$$\Delta A_t = \mathbb{E}[Y_t^2 - Y_{t-1}^2|\mathcal{F}_{t-1}] = \mathbb{E}[(Y_y - Y_{t-1})^2|\mathcal{F}_{t-1}] = \mathbb{E}[(\Delta Y_t)^2|\mathcal{F}_{t-1}]$$

so that $A_t = \sum_{s=1}^{t} \mathbb{E}[(\Delta Y_s)^2|\mathcal{F}_{s-1}]$. Prove this.

□

In the continuous–time theory, the generalization of the Doob decomposition to the Doob–Meyer Decomposition Theorem for submartingales is a deep result. The quadratic variation process associated with a submartingale is of great importance in deriving a general theory of stochastic integration.

Definition 2.1.11 If C is a previsible process, and if X is adapted (both with respect to a filtration $\mathcal{F}_n, n \in \mathbb{N}$), then the **martingale transform** of X by C is the process W given by

$$W_0 = 0$$
$$W_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) \qquad \text{if } n > 0$$

The process W is generally denoted by $C \cdot X$, and W_n by $(C \cdot X)_n$.

Thus the martingale transform of X by C is simply your winnings process on the game X using the gambling strategy C. Now comes the crunch:

Theorem 2.1.12 *(a) Suppose that X is a martingale, and that C is a bounded previsible process. Then $C \cdot X$ is a martingale.*

(b) If X is a supermartingale (submartingale), and C is a bounded non–negative *previsible process, then $C \cdot X$ is a supermartingale (submartingale).*

Proof: (a) Let $W = C \cdot X$. The fact that C is bounded and that each X_n is integrable implies that W_n is integrable as well. Then $W_{n+1} - W_n = C_{n+1}(X_{n+1} - X_n)$. Using the fact that X_n and C_{n+1} are $\mathcal{F}_n$–measurable, we see that

$$\mathbb{E}[W_{n+1} - W_n | \mathcal{F}_n] = C_{n+1}[\mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n] = 0$$

so that $\mathbb{E}[W_{n+1}|\mathcal{F}_n] = \mathbb{E}[W_n|\mathcal{F}_n] = W_n$.

The proof of (b) is left as an exercise.

$\dashv$

This theorem has the following important consequence for games of chance: You cannot find a previsible trading strategy which will turn a fair game to your advantage, i.e. which will turn a martingale into a submartingale. No matter what your strategy, your winnings process will still be a martingale.

As a final remark, note that if X is a (super–, sub–)martingale, and a is a constant, then $Y = X + a$ is a (super–, sub–)martingale, and moreover $C \cdot X = C \cdot Y$.

2.2 Stopping Times and Optional Stopping

In many games of chance, and this includes playing the stock market, one has the option to quit at any time. You may have a strategy to decide when to stop, e.g. quit if you've lost 5 times in a row, or quit if you've lost half your initial fortune. In that case, your stopping time depends on the state of the world, i.e. it is random. In the discrete framework, we can therefore regard a stopping time as a random variable $\tau : \Omega \to \mathbb{N}$. If $\tau(\omega) = n$, then you stop

after the n^{th} game if the state of the world is ω. Not all random variables $\tau : \Omega \to \mathbb{N}$ are suitable as stopping strategies, however. For example, let W_n be your winnings after the n^{th} round of a coin–tossing game. Define

$$\tau = n \qquad \text{where } W_n = \sup\{W_m : m \leq 30\}$$

This is a very desirable stopping strategy. Here's why: The strategy considers a sequence of 30 games, and stops when the winnings are at a maximum. Thus if $\tau = 23$, then W_{23} is the largest amount you will win in this state of the world. $W_{24}, W_{25}, \ldots W_{30}$ are all $\leq W_{23}$. Clearly the best thing to do is to stop at game 23. However, the problem is that by the time you reach game 23, *you don't know* whether W_{23} is the highest your winnings will ever be. This information is not available. Therefore not all positive integer–valued random variables are good stopping times. We define:

Definition 2.2.1 A map $\tau : \Omega \to \{0, 1, 2 \ldots, \infty\}$ is called a **stopping time** if

(a) $\{\tau \leq n\} = \{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n \qquad \text{for all } n \leq \infty$

or equivalently, if

(b) $\{\tau = n\} \in \mathcal{F}_n \qquad \text{for all } n \leq \infty$

Intuitively,

τ is a stopping time if and only if at time n you can decide whether $\tau = n$ or not. Whether you continue or stop depends only on the history up to, and including, time n.

Note that we include the possibility that $\tau = \infty$, i.e. that the game never stops.

Proof that (a) and (b) are equivalent: Suppose that (a) holds, i.e. that $\{\tau \leq k\} \in \mathcal{F}_k$ for all k. Then

$$\{\tau = n\} = \{\tau \leq n\} - \{\tau \leq n-1\} \in \mathcal{F}_n$$

On the other hand, if τ has property (b), i.e. if $\{\tau = k\} \in \mathcal{F}_k$ for all k, then

$$\{\tau \leq n\} = \bigcup_{k \leq n} \{\tau = k\} \in \mathcal{F}_n$$

$\dashv$

Example 2.2.2 Suppose you and an opponent play a coin tossing game, both with initial fortunes of R10.00. Let S_n denote your fortune after the 10^{th} game. Then

$$\tau = \min\{n : S_n = 0 \text{ or } S_n = 15\}$$

is clearly a stopping time (with respect to the filtration $\mathcal{F}_n = \sigma(S_0, S_1, \ldots, S_n)$): You will stop playing either when you are ruined (i.e. when $S_n = 0$), or when you've won R5.00 off your opponent.

Using the mathematical definition, we have

$$\begin{aligned}\{\tau = n\} = &\{0 < S_1 < 15\} \cap \{0 < S_2 < 15\} \cap \ldots \\ &\cdots \cap \{0 < S_{n-1} < 15\} \cap (\{S_n = 15\} \cup \{S_n = 0\})\end{aligned}$$

and each of the sets on the right belongs to $\mathcal{F}_n$. Hence so does $\{\tau = n\}$.

In this case, τ is called a *hitting time*: It is the first time the process S_n hits either 0 or 15.

$\square$

Exercise 2.2.3 Let X_n be a stochastic process adapted to a filtration $\mathcal{F}_n$, and let $B \subseteq \mathbb{R}$ be a Borel set. Show that the time of first entry into B,

$$\tau = \min\{n : X_n \in B\}$$

is a stopping time.

$\square$

Recall the following terminology: If a, b are real numbers, then

$$a \wedge b = \min\{a, b\}$$

The next exercise is important:

Exercise 2.2.4 (a) Prove that if S, T are stopping times, then so are $T \wedge S, T \vee S, T + S$.

(b) Prove that if $T_n, n \in \mathbb{N}$ are stopping times, then so are $\sup_n T_n, \inf_n T_n, \limsup_n T_n, \liminf_n T_n$, and where it exists, $\lim_n T_n$.

$\square$

Given a stochastic process X_n (e.g. your winnings in a game of chance), and a stopping time τ, we define the stopped value X_τ to be the random variable defined by

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega)$$

We also define the *stopped process* X_n^τ to be the same as X_n until the stopping time is reached, and constant with value X_τ thereafter. To be precise:

Definition 2.2.5 Let X_n be a stochastic process, and let τ be a stopping time. We define the **stopped process** X_n^τ by

$$X_n^\tau(\omega) = X_{\tau \wedge n}(\omega) = \begin{cases} X_n(\omega) \text{ if } n < \tau(\omega) \\ X_{\tau(\omega)} \text{ if } n \geq \tau(\omega) \end{cases}$$

$\square$

It is easy to see that if X_n is adapted to a filtration $\mathcal{F}_n$, then so is X_n^τ.

$\square$

In the previous section, we showed that you cannot turn a fair game to your advantage by choosing an appropriate betting strategy. Our next result shows that you cannot turn a fair game to your advantage by choosing an appropriate stopping time:

Theorem 2.2.6 (Stopped martingales are martingales)
Let τ be a stopping time.

(a) If X is a martingale, then so is the stopped process X^τ.

(b) If X is a supermartingale (submartingale), then so is X^τ.

Proof: This follows from Theorem 2.1.12. First assume that $X_0 = 0$. We will show that X^τ is the martingale transform of X with respect to a suitable strategy. Define a previsible process C by

$$C_n = \begin{cases} 0 \text{ if } \tau < n \\ 1 \text{ if } \tau \geq n \end{cases}$$

Thus $C_n = I_{\{\tau \geq n\}}$. Now $\{\tau \geq n\} = \{\tau \leq n-1\}^c$, and since τ is a stopping time, we have $\{\tau \leq n-1\} \in \mathcal{F}_{n-1}$. Hence C_n is previsible.

Now note that if we take the martingale transform of X by C, we obtain

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) = \begin{cases} X_n \text{ if } \tau \geq n \\ X_\tau \text{ if } \tau < n \end{cases}$$

and thus that $C \cdot X = X^\tau$. The result now follows from Theorem 2.1.12.

We have now proved the result for the special case where $X_0 = 0$. To prove the general result, Apply the special case to the (super–, sub–) martingale $Y_n = X_n - X_0$.

$\dashv$

This theorem immediately implies that when X is a martingale, we have $\mathbb{E}[X_n^\tau] = \mathbb{E}[X_0^\tau] = \mathbb{E}[X_0]$.

Definition 2.2.7 Suppose that τ is a stopping time on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$. The σ–algebra of *events prior* to τ, denoted $\mathcal{F}_\tau$ is the set of all events $A \in \mathcal{F}$ with the property that

$$A \cap \{\tau \leq t\} \in \mathcal{F}_t \qquad \text{for all } t \in \mathbb{T}$$

The above "definition" requires (a) of the following exercise. (This entire exercise is very important).

Exercise 2.2.8 (a) $\mathcal{F}_\tau$ is a σ–algebra.

(b) We can replace $\leq$ by $=$ in the above definition, i.e.

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = t\} \in \mathcal{F}_t \text{ for all } t \in \mathbb{T}\}$$

(c) Let X_t be an adapted process. Show that both X_τ and τ are $\mathcal{F}_\tau$–measurable.

(d) Prove that if $\sigma \leq \tau$ are stopping times, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

$\square$

How should we interpret $\mathcal{F}_\tau$? Roughly speaking,

$\mathcal{F}_\tau$ consists of all events that can be decided by time τ.

This is because $A \in \mathcal{F}_\tau$ if and only if $A \cap \{\tau \le t\}$ can be decided by time t, so if $\tau(\omega) = t$, then A is decidable at time t (i.e. at time τ). Note that though τ is random, $\mathcal{F}_\tau$ is not.

Proposition 2.2.9 *(a)* $\{\sigma \le \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$ *and* $\{\sigma = \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$*;*

(b) $A \in \mathcal{F}_\tau$ *implies* $A \cap \{\tau \le \sigma\} \in \mathcal{F}_{\sigma\wedge\tau}$ *and* $A \cap \{\tau = \sigma\} \in \mathcal{F}_{\sigma\wedge\tau}$*;*

(c) If $\tau_n \uparrow \tau < \infty$*, then* $\mathcal{F}_{\tau_n} \uparrow \mathcal{F}_\tau$*.*

Proof: (a) Note that for all n, $\{\sigma \le \tau\} \cap \{\sigma \wedge \tau = n\} = \{\sigma \le \tau\} \cap \{\sigma = n\} = \{\tau \ge n\} \cap \{\sigma = n\}$, and this set belongs to $\mathcal{F}_n$.
Similarly, $\{\tau \le \sigma\} \in \mathcal{F}_{\sigma\wedge\tau}$, and thus $\{\sigma = \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$ as well.
(b) If $A \in \mathcal{F}_\tau$, then for all n, $A \cap \{\tau \le \sigma\} \cap \{\sigma \wedge \tau = n\} = A \cap \{\tau = n\} \cap \{\sigma \le n-1\}^c \in \mathcal{F}_n$.
(c) Let $\mathcal{G} = \sigma(\bigcup_n \mathcal{F}_{\tau_n})$. We must show $\mathcal{G} = \mathcal{F}_\tau$. Now clearly $\mathcal{G} \subseteq \mathcal{F}_\tau$; hence we need only prove the reverse inclusion.

So let $A \in \mathcal{F}_\tau$. Each τ_n is $\mathcal{G}$–measurable, and thus $\tau = \lim_n \tau_n$ is $\mathcal{G}$–measurable as well. Since τ_n, τ take only integer values, since $\tau_n \uparrow \tau$, and since $\tau < \infty$, we must have $\tau(\omega) = \tau_n(\omega)$ for some n (which may depend on ω). Thus $\Omega = \bigcup_n \{\tau_n = \tau\}$. It follows that

$$A = A \cap \bigcup_n \{\tau_n = \tau\} = \bigcup_n (A \cap \{\tau_n = \tau\})$$

But by (b), since $A \in \mathcal{F}_\tau$ and $\tau_n \wedge \tau = \tau_n$, we see that $A \cap \{\tau_n = \tau\} \in \mathcal{F}_{\tau_n}$. Thus $A \in \mathcal{G}$, i.e. $\mathcal{F}_\tau \subseteq \mathcal{G}$ as well.

$\dashv$

Note: One would imagine that if τ is a stopping time and if X_n is a martingale, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$. That this is not necessarily the case is demonstrated by the following example: Let S_n be a symmetric random walk on the integers, starting at 0, and let $\tau = \min\{n : S_n = -1\}$, i.e. τ is the first time the process S_n hits -1 (and $\tau = \infty$ if it never hits -1). It is clear that S_n is a martingale with $\mathbb{E}S_n = S_0 = 0$. It is also clear that $\mathbb{E}S_\tau = -1$, because the process will stop only when it hits -1.

$\square$

Theorem 2.2.10 (Optional Sampling Theorem: Bounded Case)
Let X *be a supermartingale, and let* τ, σ *be bounded stopping times with* $\sigma \le \tau$ *a.s. Then*

$$\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \le X_\sigma \quad a.s.$$

Moreover, if X *is a martingale, then equality holds.*

Proof: Assume that $\sigma \le \tau \le N$ for some natural number N. Note that $|X_\tau| \le |X_0| + \cdots + |X_N|$, so that X_τ is integrable. The same is true for X_σ.

Next note that

$$\sum_{n=1}^{N} I_{\{\sigma \le n < \tau\}}(X_{n+1} - X_n) = X_\tau - X_\sigma$$

Now if $A \in \mathcal{F}_\sigma$, then $A \cap \{\sigma \le n\} \in \mathcal{F}_n$, and so the set

$$A_n = A \cap \{\sigma \le n < \tau\} = A \cap \{\sigma \le n\} \cap \{\tau \le n\}^c$$

belongs to $\mathcal{F}_n$. Hence

$$I_A(X_\tau - X_\sigma) = \sum_{n=1}^{N} I_{A_n}(X_{n+1} - X_n)$$

Applying the supermartingale property to $A_n \in \mathcal{F}_n$, we see that

$$\mathbb{E}[X_{n+1}; A_n] \le \mathbb{E}[X_n; A_n]$$

which implies that

$$\mathbb{E}[X_\tau; A] \le \mathbb{E}[X_\sigma; A]$$

for any $A \in \mathcal{F}_\sigma$. But then $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \le X_\sigma$, as required.

$\dashv$

Corollary 2.2.11 *Let X_n be a (super)martingale w.r.t. the filtration $\mathcal{F}_t$, and let τ_n be an increasing sequence of bounded stopping times. Then X_{τ_n} is a (super)martingale w.r.t. the filtration $\mathcal{F}_{\tau_n}$.*

Proof: Suppose that X_n is a supermartingale w.r.t. $\mathcal{F}_n$. Let $M_n = X_{\tau_n}$, $\mathcal{G}_n = \mathcal{F}_{\tau_n}$. Then if $m \le n$, we have $E[M_n|\mathcal{G}_m] = \mathbb{E}[X_{\tau_n}|\mathcal{F}_{\tau_m}] \le X_{\tau_m} = M_m$, by the Optional Sampling Theorem (bounded case), since $\tau_m \le \tau_n$ and τ_n is bounded. Clearly equality holds if X_n is a martingale.

$\dashv$

2.3 The Martingale Convergence Theorem

The main result of this section states that bounded discrete–parameter (super–, sub–) martingales converge almost surely. As usual, we work in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F}_n$.

Theorem 2.3.1 (Doob's Martingale Convergence Theorem)
Let $(X_n)_n$ be a discrete–parameter supermartingale bounded in L^1, i.e. $\sup_n \mathbb{E}(|X_n|) < \infty$. *Then there is a random variable $X_\infty \in \mathcal{L}^1$ such that*

$$X_n \to X_\infty \quad a.s. \quad as\ n \to \infty$$

Note that since $\mathbb{E}|X_\infty| < \infty$, we have $\mathbb{P}(X_\infty < \infty) = 1$, i.e. X_∞ is almost surely finite. Note also that $X_n \to X_\infty$ a.s. This does not mean that X_n converges to X_∞ in $\mathcal{L}^1$ as well. In the next section, we shall show how to extend this theorem to $\mathcal{L}^1$– and $\mathcal{L}^2$–convergence.

We need a couple of new concepts before we can tackle the proof.

Definition 2.3.2 Let X be a discrete–parameter supermartingale, and let $a < b \in \mathbb{R}$. The number $U_N(X;[a,b])(\omega)$ of *upcrossings* of $[a,b]$ is the number of times X crosses from below a to above b by time N. To be precise. $U_N(X;[a,b])(\omega)$ is the largest $k \in \mathbb{Z}^+$ for which there exist intertwined sequences s_n, t_n with

$$0 \le s_1 < t_1 < s_2 \le t_2 < \cdots < s_k < t_k \le N$$

such that

$$X_{s_i}(\omega) < a \qquad X_{t_i}(\omega) > b \qquad 1 \le i \le k$$

Also define

$$U_\infty(X;[a,b])(\omega) = \sup_N U_N(X;[a,b])(\omega)$$

We will show that $\lim_n X_n$ exists a.s. in the following manner: Suppose that $\lim_n X_n(\omega)$ does not exist. Then $\liminf_n X_n(\omega) < \limsup_n X_n(\omega)$. Thus there exist rational numbers a, b such that $\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)$, and thus $U_\infty(X;[a,b])(\omega) = \infty$. We shall show that this is possible only for a null set of ω. Thus the set $\{\liminf_n X_n \neq \limsup_n X_n\}$ has measure 0, i.e. $\lim_n X_n$ exists a.s.

We first put a bound on the number of up–crossings:

Lemma 2.3.3 (Doob's Upcrossing Lemma)
Suppose that X is a supermartingale.

$$(b-a)\mathbb{E}U_N(X;[a,b]) \le \mathbb{E}[(X_N - a)^-]$$

Proof: Regard X as a repeated game of chance, so that if you bet a stake C_n on the n^{th} game, your winnings will be $C_n(X_n - X_{n-1})$ for that game. Choose a betting strategy C as follows:

Wait until X gets below a.
Place unit stakes until X gets above b.
Wait (i.e. stop betting) until X gets below a again.
Place unit stakes until X gets above b.
etc.

To describe C mathematically, note that if $C_n = 0$ (i.e. no bets on the n^{th} game), then

$$C_{n+1} = \begin{cases} 0 \text{ if } X_n \ge a \\ 1 \text{ if } X_n < a \end{cases}$$

Similarly, if $C_n = 1$, then

$$C_{n+1} = \begin{cases} 0 \text{ if } X_n > b \\ 1 \text{ if } X_n \le b \end{cases}$$

It follows that we can define C_n inductively by

$$C_1 = I_{\{X_0 < a\}}$$
$$C_{n+1} = I_{\{C_n = 0\}} I_{\{X_n < a\}} + I_{\{C_n = 1\}} I_{\{X_n \le b\}}$$

Since C_{n+1} is defined in terms of the $\mathcal{F}_n$–measurable functions C_n, X_n, it follows that C_n is previsible.

Let Y_n be the total winnings until time n, i.e. $Y_n = (C \cdot X)_n$.

Now note that the total winnings by time N must satisfy

$$Y_N(\omega) \geq (b-a)U_N(X;[a,b])(\omega) - [X_N(\omega) - a]^-$$

This is because every upcrossing contributes at least $(b-a)$ to the total winnings. The final term, $[X_N - a]^-$, takes into account that we may be placing bets on the last stretch to time N. It is clear that our losses on this stretch cannot exceed $[X_N - a]^-$.

Since X is a supermartingale and C is previsible and non–negative, the martingale transform $C \cdot X$ is also a supermartingale. Thus

$$\mathbb{E}Y_N = \mathbb{E}(C \cdot X)_N \leq \mathbb{E}(C \cdot X)_0 = 0$$

from which the desired inequality follows immediately.

$\dashv$

Corollary 2.3.4 *Let X be a supermartingale bounded in $\mathcal{L}^1$, and let $a < b \in \mathbb{R}$. Then*

$$(b-a)\mathbb{E}U_\infty(X;[a,b]) \leq |a| + \sup_n \mathbb{E}|X_n| < \infty$$

so that

$$\mathbb{P}(U_\infty(X;[a,b]) < \infty) = 1$$

Proof: Since the U_N are non–negative and increasing, the Monotone Convergence Theorem implies that $(b-a)\mathbb{E}U_N \to (b-a)\mathbb{E}U_\infty$ as $N \to \infty$. Now, using the triangle inequality,

$$\begin{aligned}(b-a)\lim_N \mathbb{E}U_N &\leq \sup_N \mathbb{E}[X_N - a]^- \\ &\leq \sup_N \mathbb{E}|X_N| + |a|\end{aligned}$$

from which the required inequality follows. The second result is a trivial consequence of the first.

$\dashv$

We can now prove the Martingale Convergence Theorem:

Proof of Martingale Convergence Theorem: We want to prove that $X_n(\omega)$ converges almost surely to some finite limit $X_\infty(\omega)$. Now recall that $\liminf_n X_n$ and $\limsup_n X_n$ always exist, but that $\lim_n X_n$ only exists if $\liminf_n X_n = \limsup_n X_n$. For concreteness' sake, *define* $X_\infty(\omega) = \limsup_n X_n(\omega)$. Let

$$\begin{aligned}\Lambda &= \{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty,\infty]\} \\ &= \{\omega : \liminf_n X_n(\omega) < \limsup_n(\omega)\} \\ &= \bigcup_{\{a<b\in\mathbb{Q}\}} \{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\} \\ &= \bigcup_{\{a<b\in\mathbb{Q}\}} \Lambda_{a,b}\end{aligned}$$

where the last equality defines $\Lambda_{a,b}$ in the obvious way. Now if

$$\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)$$

then $X_n(\omega)$ must go below a infinitely many times, and also go above b infinitely many times. It follows that $\Lambda_{a,b} \subseteq \{\omega : U_\infty(X;[a,b])(\omega) = \infty\}$. The preceding corollary assures us that $\mathbb{P}(U_\infty(X;[a,b]) = \infty) = 0$, however, and thus that $\mathbb{P}(\Lambda_{a,b}) = 0$ for every $a < b \in \mathbb{Q}$. Since Λ is just the *countable* union of the $\Lambda_{a,b}$, and since the countable union of sets of measure zero itself has measure zero, it follows that $\mathbb{P}(\Lambda) = 0$, and thus that $X_\infty = \lim_n X_n$ exists almost surely, though the limit may be $\pm\infty$. We must still prove, therefore, that $X_\infty \in \mathcal{L}^1$.

By Fatou's Lemma, we see that

$$\mathbb{E}(|X_\infty|) = \mathbb{E}(\liminf_n |X_n|) \leq \liminf_n \mathbb{E}|X_n| \leq \sup_n \mathbb{E}|X_n| < \infty$$

where the last inequality follows because we are assuming that the sequence X_n is bounded in $\mathcal{L}^1$. Hence $X_\infty \in \mathcal{L}^1$, so that $\mathbb{P}(X_\infty \text{ is finite}) = 1$, as required.

$\dashv$

Example 2.3.5 Suppose that a gambler, starting with an initial fortune $X_0 \in \mathbb{N}$, plays repeated rounds of a fair game. The gambler will play until he is ruined (if ever). Also assume that, while the gambler is playing, 1 unit is won or lost on each game.

Let X_n be the gambler's fortune after n games, and let $\tau = \inf\{n : X_n = 0\}$ be the time of ruin. Then

$$|X_{n+1} - X_n| = 1 \quad \text{if } n < \tau$$

and

$$|X_{n+1} - X_n| = 0 \quad \text{if } n \geq \tau$$

Now $X_n \geq 0$ for each n, and $\mathbb{E}X_n = X_0$, because $(X_n)_n$ is a martingale. It follows that $(X_n)_n$ is $\mathcal{L}^1$–bounded, and thus that there is $X_\infty \in \mathcal{L}^1$ such that $X_n \to X_\infty$ a.s. Thus, almost surely, $(X_n)_n$ is a Cauchy sequence.

Now let $0 < \varepsilon < 1$. Then, almost surely, there is $N \in \mathbb{N}$ such that $n \geq N$ implies $|X_{n+1} - X_n| < \varepsilon$. But then clearly $|X_{n+1} - X_n| = 0$, so that $\tau \leq N$. Hence, almost surely, τ is finite, i.e. the gambler will eventually be ruined, with probability 1.

$\square$

The Martingale Convergence Theorem states that any $\mathcal{L}^1$–bounded martingale converges almost surely. However, we are frequently interested in other types of convergence, e.g. $\mathcal{L}^p$ convergence. The next section deals with this.

We end this section with one more useful convergence result:

Proposition 2.3.6 *Suppose that $X_n \to X_\infty$ in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ for some $p \in [1, \infty)$. Let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$. Then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X_\infty|\mathcal{G}]$ in $\mathcal{L}^p$ as well.*

Proof: By Jensen's inequality,

$$\mathbb{E}\,|\mathbb{E}[X_\infty|\mathcal{G}] - \mathbb{E}[X_n|\mathcal{G}]|^p \leq \mathbb{E}|X_\infty - X_n|^p \to 0$$

$\dashv$

2.4 Uniformly Integrable Martingales

2.4.1 Uniform Integrability

Work in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.4.1 (1) A set $\{X_i : i \in I\}$ of random variables is said to be L^1–bounded if $\sup_{i\in I} \mathbb{E}|X_i| < \infty$, i.e. if there is a $K < \infty$ such that $\mathbb{E}|X_i| < K$ for all $i \in I$.

(2) A set $\{X_i : i \in I\}$ of random variables is said to be *uniformly* $\mathbb{P}$*–continuous* (u-$\mathbb{P}$-c for short) if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that whenever $\mathbb{P}(F) < \delta$, then

$$\mathbb{E}[|X|; F] < \varepsilon \qquad \text{for all } X \in \mathcal{X}$$

i.e. if and only if

$$\sup_{X\in\mathcal{X}} \mathbb{E}[|X|; F] \to 0 \text{ as } \mathbb{P}(F) \to 0$$

(3) A set $\{X_i : i \in I\}$ of random variables is said to be *uniformly integrable* (or UI) if

$$\lim_{K\to\infty} \sup_{i\in I} \mathbb{E}[|X_i|; |X_i| > K] = 0$$

i.e. if for every $\varepsilon > 0$ there is a K such that

$$\mathbb{E}[|X_i|; |X_i| > K] < \varepsilon \qquad \text{for all } i \in I$$

We say that a discrete– or continuous–parameter stochastic process X_t is UI if and only if the collection of its component random variables $\{X_t\}_t$ is UI.

Exercise 2.4.2 Suppose that $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. We show that singleton $\{X\}$ is both UI and u-$\mathbb{P}$-c.

(a) We first show that $\{X\}$ is u-$\mathbb{P}$-c. Suppose not. Explain why there is an $\varepsilon > 0$ and a sequence $F_n \in \mathcal{F}$ such that

$$\mathbb{P}(F_n) < 2^{-n} \qquad \text{but} \qquad \mathbb{E}[|X|; F_n] > \varepsilon$$

(b) Now define $F := (F_n, i.o.)$. Apply a Fatou Lemma to conclude that

$$\mathbb{E}[|X|; F] \geq \limsup_n \mathbb{E}[|X|; F_n] \geq \varepsilon$$

(c) Apply a Borel–Cantelli Lemma to show

$$\mathbb{P}(F) = \mathbb{P}(F_n, \text{i.o.}) = 0$$

and explain why we have obtained a contradiction. This proves that $\{X\}$ is u-$\mathbb{P}$-c.

(d) Next, we show that $\{X\}$ is UI. Let $\varepsilon > 0$, and choose $\delta > 0$ such that $\mathbb{E}[|X|; F] < \varepsilon$ whenever $\mathbb{P}(F) < \delta$. Why can we do this?

(e) Show that

$$\mathbb{P}(|X| > K) \leq \frac{1}{K}\mathbb{E}|X| < \infty$$

(f) Take $K > \frac{\mathbb{E}|X|}{\delta}$, and show $\mathbb{E}[|X|; |X| > K] < \varepsilon$, as required.

□

Exercises 2.4.3 (a) $\{X_i : i \in I\}$ is UI if and only if $\{|X_i| : i \in I\}$ is UI.

(b) Every UI family is bounded in $\mathcal{L}^1$.

(c) Not every $\mathcal{L}^1$–bounded family is UI: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the usual probability space on the unit interval $[0,1]$, together with Lebesgue measure. Let

$$F_n = [0, 1/n] \qquad X_n = nI_{F_n}$$

for $n \in \mathbb{N}$. Show that $\{X_n : n \in \mathbb{N}\}$ is bounded in $\mathcal{L}^1$, but that $\{X_n : i \in \mathbb{N}\}$ is not UI.

(d) Suppose that $\{X_i : i \in I\}$ is a family of random variables dominated by some $Y \in \mathcal{L}^1$ (i.e. $|X_i| \leq Y$ a.s. for all $i \in I$). Show that $\mathcal{X}$ is UI.

(e) Show that any finite family of integrable random variables is UI.

(f) Show that if $\mathcal{X}, \mathcal{Y}$ are two UI families, then the families $\mathcal{X} \cup \mathcal{Y}$ and $\mathcal{X} + \mathcal{Y} = \{X + Y : X \in \mathcal{X}, Y \in \mathcal{Y}\}$ are also UI.

(g) Show that if $\mathcal{X}$ is a family of identically distributed integrable random variables, then $\mathcal{X}$ is UI.

□

Theorem 2.4.4 *$\{X_i : i \in I\}$ is UI iff it is L^1–bounded and u-$\mathbb{P}$-c.*

Proof: ($\Rightarrow$): Suppose $\{X_i : i \in I\}$ is L^1–bounded and u-$\mathbb{P}$-c. Choose $M < \infty$ such that $\mathbb{E}|X_i| \leq M$ for all $i \in I$ (by L^1–boundedness). For $\varepsilon > 0$, choose a $\delta > 0$ as in the definition of uniform-$\mathbb{P}$–continuity. Observe that $\mathbb{K}P(|X_i| > K) \leq \mathbb{E}[|X_i|; |X_i| > K] \leq M$, so that if $K > M/\delta$, then $\mathbb{P}(|X_i| > K) < \delta)$, from which $\mathbb{E}[|X_i|; |X_i| > K] < \varepsilon$. Since the definition of K does not depend on $i \in I$, we see that $\{X_i : i \in I\}$ is UI.

($\Leftarrow$): Now suppose that $\{X_i : i \in I\}$ is UI. For $\varepsilon > 0$, choose K such that $\mathbb{E}[|X_i|; |X_i| > K] < \varepsilon/2$ for all $i \in I$. Observe first that

$$\mathbb{E}|X_i| = \mathbb{E}[|X_i|; |X_i| \leq K] + \mathbb{E}[|X_i|; |X_i| > K] \leq K + \tfrac{\varepsilon}{2} \qquad \text{for all } i \in I$$

which proves that $\{X_i : i \in I\}$ is L^1–bounded. Now if F is a measurable set with $\mathbb{P}(F) \leq \frac{\varepsilon}{2K}$, we have

$$\mathbb{E}[|X_i|; F] = \mathbb{E}[|X_i|; F \cap \{|X_i| < K\}] + \mathbb{E}[|X_i|; F \cap \{|X_i| \geq K\}] < K\mathbb{P}(F) + \varepsilon/2 \leq \varepsilon$$

so that $\{X_i : i \in I\}$ is also u-$\mathbb{P}$-c.

⊣

Our main result is the followinG: L^1–convergence is precisely the intersection of uniform integrability with convergence in probability.

Theorem 2.4.5 *If $X_n, X \in \mathcal{L}^1$, then $X_n \xrightarrow{L^1} X$ if and only if $\{X_n : n \in \mathbb{N}\}$ is UI* and *$X_n \xrightarrow{\mathbb{P}} X$.*

Proof: ($\Rightarrow$): We know that if $X_n \xrightarrow{L^1} X$, then $\{X_n : n \in \mathbb{N}\}$ is L^1–bounded and $X_n \xrightarrow{\mathbb{P}} X$. It therefore remains to show that $\{X_n : n \in \mathbb{N}\}$ is u-$\mathbb{P}$-c. So let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ so that $n \geq N$ implies $||X_n - X||_1 = \mathbb{E}|X_n - X| < \varepsilon/2$. Choose $\delta > 0$ so that $\mathbb{P}(F) < \delta$ implies $\mathbb{E}[|X|; F] < \varepsilon/2$ (cf. Exercise 2.4.2), and decrease δ if necessary so that also $\mathbb{P}(F) < \delta)$ implies $\mathbb{E}[|X_n|; F] < \varepsilon/2$ for $n = 1, \ldots, N-1$. Then if $n \geq N$ we also have

$$\mathbb{E}[|X_n|; F] \leq \mathbb{E}[|X_n - X|; F] + \mathbb{E}[|X|; F] < \varepsilon/2 + \varepsilon/2$$

which shows that $\{X_n : n \in \mathbb{N}\}$ is u-$\mathbb{P}$-c.

($\Leftarrow$): Recall that we proved the following fact about convergence in probability: *Suppose that $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a function which is increasing, and strictly increasing on some interval $(0, a)$, is bounded, continuous, and satisfies $f(0) = 0$. Then $X_n \xrightarrow{\mathbb{P}} X$ iff $\mathbb{E}f|X_n - X| \to 0$.* Suppose now that $\{X_n\}_n$ is UI, and that $X_n \xrightarrow{\mathbb{P}} X$. Fix $\varepsilon > 0$. Note that $\{X_n - X : n \in \mathbb{N}\}$ is UI also (cf. Exercises 2.4.2,2.4.3 (f)). First, pick K such that $\mathbb{E}[|X_n - X|; |X_n - X| \geq K] < \varepsilon/2$ for all $n \in \mathbb{N}$. Now set $f(x) := \wedge K$, and note that f has the properties required to determine convergence in probability. Hence $\mathbb{E}[|X_n - X| \wedge K] \to 0$. Choose $N \in \mathbb{N}$ sufficiently large that $\mathbb{E}[X_n - X| \wedge K] < \varepsilon/2$ whenenever $n \geq N$. Then for $n \geq N$, we have

$$\begin{aligned} ||X_n - X||_1 = \mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X|; |X_n - X| \leq K] + \mathbb{E}[X_n - X|; |X_n - X| > K] \\ &\leq \mathbb{E}[|X_n - X| \wedge K] + \mathbb{E}[X_n - X|; |X_n - X| > K] < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} \end{aligned}$$

so that $||X_n - X||_1 \to 0$, as required.

$\dashv$

We continue with a few more results about uniform integrability that will come in handy:

We have already seen that every UI family is bounded in L^1, but that not every L^1–bounded family is UI (by Exercise 2.4.3). Nevertheless, being UI is only just stronger than being bounded in L^1, as the following proposition makes clear:

Proposition 2.4.6 *If $p > 1$ and $\{X_i : i \in I\}$ is a family of random variables which is bounded in L^p, then $\{X_i\}_i$ is* UI.

Exercise 2.4.7 We prove the preceding Proposition:

(a) Explain why there is $B \in \mathbb{R}$ such that $\mathbb{E}|X_i|^p \leq B$ for all $i \in I$.

(b) Let $K > 0$. Show that if $x \geq K$, then $x = x^{1-p}x^p \leq K^{1-p}x^p$

(c) Deduce that
$$\mathbb{E}[|X_i|; |X_i| > K] \leq K^{1-p}\mathbb{E}[|X_i|^p; |X_i| > K] \leq K^{1-p}B$$
for all $i \in I$.

(d) Now given $\varepsilon > 0$, choose K sufficiently large so that $K^{1-p}B < \varepsilon$. Show that we will have $\mathbb{E}[|X_i|; |X_i| > K] < \varepsilon$ for all $i \in I$, so that $\{X_i : i \in I\}$ is UI, as claimed.

$\square$

Here is an important source of UI martingales:

Theorem 2.4.8 *Suppose that $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, and that $\{\mathcal{F}_i : i \in I\}$ is a family of sub-σ-algebras of $\mathcal{F}$. Then the set*

$$\Big\{\mathbb{E}[X|\mathcal{F}_i] : i \in I\Big\}$$

is UI.

The proof is an exercise:

Exercise 2.4.9 Prove the preceding proposition.
[Hint: Let $\varepsilon > 0$, and choose $\delta > 0$ such that $\mathbb{P}(F) < \delta \Rightarrow \mathbb{E}(|X|; F) < \varepsilon$. (Why does δ exist?) Choose K such that $K^{-1}\mathbb{E}|X| < \delta$. If $\mathcal{F}_i \subseteq \mathcal{F}$, let $X_i := \mathbb{E}[X|\mathcal{F}_i]$, and use Jensen's inequality combined with Markov's inequality to show that

$$K\mathbb{P}(|X_i| > K) \leq \mathbb{E}|X|$$

Deduce from the definition of conditional expectation that $\mathbb{E}[|X_i|; |X_i| > K] < \varepsilon$.]

□

Remarks 2.4.10 The importance of the notion of uniform integrability becomes clear when we consider topology: Uniform integrability is equivalent to relative sequential compactness in L^1 equipped with its weak (i.e. $\sigma(L_1, L_\infty)$–) topology. We will not need this fact, so do not prove it here.

□

2.4.2 UI Martingales

Theorem 2.4.11 *(a) Suppose that X is a supermartingale. Then X is UI if and only if there is a random variable X_∞ such that $X_n \to X_\infty$ a.s.* and *in $\mathcal{L}^1$. We then have $X_n \geq \mathbb{E}[X_\infty|\mathcal{F}_n]$.*

(b) Moreover, if X is a UI martingale,then

$$X_n = \mathbb{E}[X_\infty|\mathcal{F}_n] \quad \text{a.s.}$$

Proof: (a) Suppose that $(X_n)_n$ is a UI supermartingale. Then X_n is bounded in $\mathcal{L}^1$, and thus there is a random variable X_∞ such that $X_n \to X_\infty$ a.s., by the Martingale Convergence Theorem. But then $X_n \to X_\infty$ in probability. It follows that $X_n \to X_\infty$ in $\mathcal{L}^1$, by Theorem 2.4.5.

Conversely, if $X_n \to X_\infty$ in L^1, then $(X_n)_n$ is UI, again by Theorem 2.4.5.

We now show that $\mathbb{E}[X_\infty|\mathcal{F}_n] = X_n$ a.s. Suppose that $F \in \mathcal{F}_n$, then

$$\mathbb{E}[X_n; F] \geq \mathbb{E}[X_m; F] \qquad \text{for all } m \geq n$$

(This is just the supermartingale property.) But

$$|\mathbb{E}[X_m; F] - \mathbb{E}[X_\infty; F]| \leq \mathbb{E}[|X_m - X_\infty|; F] \to 0 \quad \text{as } m \to \infty$$

because $X_m \to X_\infty$ in $\mathcal{L}^1$. Thus, letting $m \to \infty$, we get

$$\mathbb{E}[X_m; F] \to \mathbb{E}[X_\infty; F]$$

But for $m \geq n$ we have $\mathbb{E}[X_m; F] \leq \mathbb{E}[X_n; F]$, and thus $\mathbb{E}[X_\infty; F] \leq \mathbb{E}[X_n; F]$, as required.

(b) follows from the observation that if X is a martingale, then we can replace the inequality signs by $=$ in the above.

⊣

Remarks 2.4.12 We have seen in Theorem 2.4.8 that applying conditional expectations to an integrable random variable produces uniformly integrability. By Theorems 2.4.8 and 2.4.11 it follows straight away that *all* UI martingales are obtained by applying conditional expectations:

A martingale M is UI if and only if there is a random variable M_∞ such that $M_n \to M_\infty$ a.s. and in $\mathcal{L}^1$, and

$$M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$$

□

2.4.3 Optional Sampling of UI Supermartingales

Let M_n be a UI martingale with respect to some filtration $\mathcal{F}_n$, and let $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.

Theorem 2.4.13 (Doob's Optional Sampling Theorem) *Let $0 \le \sigma \le \tau \le \infty$ be stopping times. and suppose that M is a UI martingale. Then*

$$\mathbb{E}[M_\tau | \mathcal{F}_\sigma] = M_\sigma \ \textit{a.s.}$$

Proof: We know that there is M_∞ such that $M_n \to M_\infty$ a.s. and in L^1 and that $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ for all n. To prove the theorem, it suffices to show that $\mathbb{E}[M_\infty | \mathcal{F}_\tau] = M_\tau$, for then

$$\mathbb{E}[M_\tau | \mathcal{F}_\sigma] = \mathbb{E}[M_\infty | \mathcal{F}_\tau | \mathcal{F}_\sigma] = \mathbb{E}[M_\infty | \mathcal{F}_\sigma] = M_\sigma$$

Now If $F \in \mathcal{F}_\tau$, then $F = \bigcup_n F \cap \{\tau = n\}$, and each $F \cap \{\tau = n\} \in \mathcal{F}_n$. Hence

$$\begin{aligned}
\mathbb{E}[M_\infty; F] &= \sum_n \mathbb{E}[M_\infty; F \cap \{\tau = n\} \\
&= \sum_n \mathbb{E}[M_n; F \cap \{\tau = n\} \\
&= \sum_n \mathbb{E}[M_\tau; F \cap \{\tau = n\}] \\
&= \mathbb{E}[M\tau; F]
\end{aligned}$$

⊣

We can now give another characterization of UI martingales. Let M be an adapted process. We define M_∞ by

$$M_\infty(\omega) = \begin{cases} \lim\limits_{n\to\infty} M_n(\omega) & \text{if this limit exists} \\ 0 & \text{otherwise} \end{cases}$$

Note that if $M_n \to X$ a.s., then $X = M_\infty$ a.s.

Theorem 2.4.14 *Suppose that M is an adapted process. Then the following are equivalent:*

(a) M is a UI *martingale;*

(b) There is $c \in \mathbb{R}$ such that for every stopping time $\tau \le \infty$, we have

$$\mathbb{E}\|M_\tau|] < \infty \quad \textit{and} \quad \mathbb{E}[M_\tau] = c$$

Proof: (a) $\Rightarrow$ (b): If M is a UI martingale, then we know that $M_n \to M_\infty$ a.s. and in $\mathcal{L}^1$. By applying the Optional Sampling Theorem to the stopping times $0 \leq \tau$, we have $\mathbb{E}[M_\tau] = \mathbb{E}M_0 = c$ for all stopping times τ. Moreover, if we apply the Optional Sampling Theorem to the stopping times $\tau \leq \infty$, we see that $M_\tau = \mathbb{E}[M_\infty|\mathcal{F}_\tau]$. Now $M_\infty \in \mathcal{L}^1$, and $\mathbb{E}[|M_\tau|] \leq \mathbb{E}[|M_\infty|] < \infty$, by Jensen's inequality. Thus $M_\tau \in \mathcal{L}^1$ for every stopping time τ.
(b) $\Rightarrow$ (a): Note that every constant time $\tau = n$ is a stopping time, and thus we have $M_n \in \mathcal{L}^1$ with $\mathbb{E}[M_n] = c$ for all $n \leq \infty$. (Here M_∞ is defined in the way described just before the statement of the theorem.) This *suggests* a martingale property (but is still a long way from proving it). Now let $F \in \mathcal{F}_n$, and define the random time τ by

$$\tau(\omega) = \begin{cases} n \text{ if } \omega \in F \\ \infty \text{ if } \omega \in F^c \end{cases}$$

It is clear that τ is a stopping time, and thus

$$\begin{aligned} c = \mathbb{E}[M_\tau] = \mathbb{E}[M_n; F] + \mathbb{E}[M_\infty; F^c] \\ c = \mathbb{E}[M_\infty] = \mathbb{E}[M_\infty; F] + \mathbb{E}[M_\infty; F^c] \end{aligned}$$

which clearly implies that $\mathbb{E}[M_n; F] = \mathbb{E}[M_\infty; F]$. It follows that

$$M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]$$

and thus that M is a UI martingale, by Theorem 2.4.8

$\dashv$

2.5 Upwards and Downwards

Theorem 2.5.1 (Lévy's Upward Theorem)
Suppose that $\zeta \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, and that $\mathcal{F}_n$ is a filtration, $\mathcal{F}_n \subseteq \mathcal{F}$. Define $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$, and define $M_n = \mathbb{E}[\zeta|\mathcal{F}_n]$ for $n \leq \infty$. Then M is a UI *martingale and $M_n \to M_\infty$ a.s. and in $\mathcal{L}^1$.*

Proof: That M is a martingale follows trivially from the Tower Property, and that M is UI follows from Theorem 2.4.8. Hence there is η such that $M_n \to \eta$ a.s. and in $\mathcal{L}^1$, so we must just show that $\eta = M_\infty$ a.s. But if $F \in \mathcal{F}_n$, then

$$\mathbb{E}[\zeta; F] = \mathbb{E}[M_m; F] \text{ for all } m \geq n$$

by definition of conditional expectation. Since $M_m \to \eta$ in $\mathcal{L}^1$, we must have $\mathbb{E}[\eta; F] = \mathbb{E}[\zeta; F]$ for all $F \in \mathcal{F}_n$. Since this is true for all n, we have

$$\mathbb{E}[\eta; F] = \mathbb{E}[\zeta; F] \text{ for all } F \in \bigcup_n \mathcal{F}_n$$

But $\bigcup_n \mathcal{F}_n$ is a π–system that generates $\mathcal{F}_\infty$, and thus $\eta = \mathbb{E}[\zeta|\mathcal{F}_\infty] = M_\infty$ a.s.

$\square$

There is also a downwards version of the preceding theorem, obtained by going backwards in time, which will play an important part in the continuous–parameter theory. This necessitates the introduction of *reversed martingales.*

Note that we can define the notion of martingale w.r.t. any partially ordered index set $(P, \leq)$ as follows: A P–filtration is a set of P–indexed σ–algebras satisfying

$$\mathcal{F}_p \subseteq \mathcal{F}_q \qquad \text{whenever } p \leq q \text{ in } P$$

and an adapted P–indexed family of integrable random variables X_p is called a P–supermartingale if and only if

$$\mathbb{E}[X_q|\mathcal{F}_p] \leq X_p \qquad \text{whenever } p \leq q \text{ in } P$$

Note that this definition makes sense even if P is not a total ordering.

Now consider the set $\mathbb{N}$ of non–negative integers together with the reverse ordering $\precsim$ defined by

$$n \precsim m \qquad \text{iff} \qquad m \leq n$$

A $(\mathbb{N}, \precsim)$–supermartingale is called a *reversed supermartingale.* Thus if X_n is a reversed supermartingale with respect to a filtration $\mathcal{F}_n$, then each X_n is integrable and $\mathcal{F}_n$–measurable, and

$$\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{F}_3 \dots \qquad \text{and} \qquad \mathbb{E}[X_m|\mathcal{F}_n] \leq X_n \quad \text{if } m \leq n$$

In particular, $\mathbb{E}[X_n|\mathcal{F}_{n+1}] \leq X_{n+1}$ and $\mathbb{E}[X_0|\mathcal{F}_n] \leq X_n$

Theorem 2.5.2 (Lévy–Doob Downward Theorem)
Let $(\mathcal{F}_n : n \in \mathbb{N})$ be a decreasing sequence of σ–algebras, i.e.

$$\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{F}_3 \dots$$

and define $\mathcal{F}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{F}_n$ Let $X = (X_n : n \in \mathbb{N})$ be a reversed supermartingale w.r.t. $(\mathcal{F}_n : n \in \mathbb{N})$, so that

$$\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m \qquad \textit{for } n \leq m$$

Finally, assume that the family X_n has $\lim_n \mathbb{E}X_n < \infty$.
Then X_n is UI, *and the limit*

$$X_\infty = \lim_{n\to\infty} X_n$$

exists a.s. and in $\mathcal{L}^1$. Moreover, we have

$$\mathbb{E}[X_n|\mathcal{F}_\infty] \leq X_\infty \quad \text{a.s.}$$

with equality if X_n is a reversed martingale.

Proof: The existence of the a.s. limit $X_{-\infty} = \lim\limits_{n\to-\infty} X_n$ follows from the Upcrossing Lemma, just as for the ordinary Supermartingale Convergence Theorem: Note that if $X_0, \dots, X_N$ is a reversed supermartingale, then the reversed sequence $Y_n = X_{N-n}$ is an ordinary supermartingale. By the Upcrossing Lemma, we therefore have $U_N(Y, [a, b]) \leq \mathbb{E}[(Y_N - a)^-]$. Clearly Y and X have the same number of upcrossings, however, and thus $U_N(X; [a, b]) \leq \mathbb{E}[(X_0 - a)^-]$. From here, it is straightforward to prove the a.s. convergence of X_n.

Once we've proved that X is UI, it will follow from Theorem 2.4.5 that X_n converges in $\mathcal{L}^1$ as well. A quick perusal of the proof of Theorem 2.4.11 ought to convince you that in that case we also have $\mathbb{E}[X_m|\mathcal{F}_\infty] \leq X_\infty$ a.s., with equality if X is a martingale.

Our task is therefore to prove that X is UI. Let $\varepsilon > 0$, and note that $(\mathbb{E}X_n)_n$ is an increasing bounded sequence (by hypothesis). Since a monotone bounded sequence is convergent, it is Cauchy, and thus we may choose $N \in \mathbb{N}$ such that $n \geq N$ implies $\mathbb{E}X_n - \mathbb{E}X_N < \frac{\varepsilon}{2}$. Then for $K > 0$, we have

$$\begin{aligned}
\mathbb{E}[|X_n| : |X_n| > K] &= -\mathbb{E}[X_n; X_n < -K] + \mathbb{E}[X_n] - \mathbb{E}[X_n; X_n \leq K] \\
&\leq -\mathbb{E}[X_N; X_n < K] + \mathbb{E}[X_N] + \frac{\varepsilon}{2} - \mathbb{E}[X_N; X_n \leq K] \\
&= \mathbb{E}[-X_N; X_n < -K] + \mathbb{E}[X_N; X_n > K] + \frac{\varepsilon}{2} \\
&\leq \mathbb{E}[|X_N|; |X_n| \geq K] + \frac{\varepsilon}{2}
\end{aligned}$$

But $\mathbb{P}(|X_n| > K) \to 0$ as $n \to \infty$: For let $L =\uparrow \lim_n \mathbb{E}X_n < \infty$. By Jensen's inequality, X_n^- is a reversed *sub*martingale (because $x^- = \max\{-x, 0\}$ is clearly convex and decreasing). Thus if $K > 0$, we have

$$K\mathbb{P}(|X_n| > K) \leq \mathbb{E}|X_n| = \mathbb{E}X_n + 2\mathbb{E}X_n^- \leq L + 2\mathbb{E}X_0^-$$

It follows that $\sup_n \mathbb{P}(|X_n| > K) \to 0$ as $K \to \infty$.

Hence by picking a sufficiently large K, we can ensure that $\mathbb{E}[|X_n|; |X_n| > K] < \varepsilon$ for all $n \geq N$. If necessary, we can enlarge K even more to ensure that $\mathbb{E}[|X_n|; |X_n| > K] < \varepsilon$ for all $n < N$ as well.

This proves that X is UI, as required.

⊣

It may be useful to put the upwards and downwards theorems together. Let $\mathcal{F}_n \uparrow \mathcal{G}$ mean "$\mathcal{F}_n$ is increasing and $\mathcal{G} = \sigma(\bigcup_n \mathcal{F}_n)$". Similarly, $\mathcal{F}_n \downarrow \mathcal{G}$ abbreviates "$\mathcal{F}_n$ is decreasing, and $\mathcal{G} = \bigcap_n \mathcal{F}_n$."

Theorem 2.5.3 *Suppose that $\mathcal{F}_n$ is a sequence of σ–algebras and that ζ is an integrable random variable. Define $X_n = \mathbb{E}[\zeta|\mathcal{F}_n]$.*

(a) If $\mathcal{F}_n \uparrow \mathcal{G}$, then $X_n \to \mathbb{E}[\zeta|\mathcal{G}]$ a.s. and in $\mathcal{L}^1$. Moreover, X_n is a UI martingale.

(b) If $\mathcal{F}_n \downarrow \mathcal{G}$, then $X_n \to \mathbb{E}[\zeta|\mathcal{G}]$ a.s. and in $\mathcal{L}^1$. Moreover, X_n is a UI reversed martingale.

With the Lévy–Doob results, it is now possible to prove a stronger version of the Lebesgue Dominated Convergence Theorem. First note the following useful fact:

Proposition 2.5.4 *(a) If $\mathcal{F}_n \uparrow \mathcal{G}$ and $X_n \to X$ in $\mathcal{L}^1$, then $\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}]$ in $\mathcal{L}^1$.*

(b) If $\mathcal{F}_n \downarrow \mathcal{G}$ and $X_n \to X$ in $\mathcal{L}^1$, then $\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}]$ in $\mathcal{L}^1$.

Proof: To prove (a), we must show that $||\mathbb{E}[X_n|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]||_1 = \mathbb{E}|\mathbb{E}[X_n|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]| \to 0$ as $n \to \infty$. Now by the triangle inequality,

$$\begin{aligned}||\mathbb{E}[X_n|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]||_1 &\leq ||\mathbb{E}[X_n|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{F}_n]||_1 + ||\mathbb{E}[X|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]||_1 \\ &\leq ||X_n - X||_1 + ||\mathbb{E}[X|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]||_1\end{aligned}$$

But $||X_n - X||_1 \to 0$ by hypothesis, and $||\mathbb{E}[X|\mathcal{F}_n]-\mathbb{E}[X|\mathcal{G}]||_1 \to 0$ by Lévy's Upward Theorem.
(b) is proved in the same way, this time invoking the Downward Theorem.

⊣

Theorem 2.5.5 (Strong Dominated Convergence Theorem)
Let $\mathcal{F}_n, \mathcal{G}$ be σ–algebras, and let X_n, X, Z be integrable random variables. Assume further that $X_n \to X$ a.s., and that each $|X_n| \leq Z$ a.s. Then

(a) If $\mathcal{F}_n \uparrow \mathcal{G}$, then $\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}]$ a.s. and in $\mathcal{L}^1$

(b) If $\mathcal{F}_n \downarrow \mathcal{G}$, then $\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}]$ a.s. and in $\mathcal{L}^1$.

Proof: By the Lebesgue Dominated Convergence Theorem, we have $X_n \to X$ in $\mathcal{L}^1$, and thus the $\mathcal{L}^1$–convergence of conditional expectations follows from the previous proposition. We need therefore only show that a.s. convergence holds.

We only prove (a), the proof of (b) being very similar.
Define $W_n = \sup_{k \geq n}|X_k - X|$. Then $W_n \in \mathcal{L}^1$, because $W_n \leq 2Z$. Since $X_n \to X$ a.s., we also must have $W_n \downarrow 0$ a.s. Now fix $N \geq 1$. If $n \geq N$, then $|X_n - X| \leq W_n \leq W_N$, which implies $\mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_n]$. By the Upward Theorem, we must have $\mathbb{E}[W_N|\mathcal{F}_n] \to \mathbb{E}[W_N|\mathcal{G}]$ as $n \to \infty$, a.s. and in $\mathcal{L}^1$. It follows that

$$\limsup_n \mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \lim_n \mathbb{E}[W_N|\mathcal{F}_n] = \mathbb{E}[W_N|\mathcal{G}]$$

Now let $N \to \infty$. By the Lebesgue Dominated Convergence Theorem for conditional expectations, since $W_N \leq 2Z$ and $W_N \downarrow 0$ a.s., we also have $\mathbb{E}[W_N|\mathcal{G}] \downarrow 0$ a.s. It follows that $\limsup_n \mathbb{E}[|X_n - X||\mathcal{F}_n] = 0$ a.s., and thus

$$|\mathbb{E}[X_n|\mathcal{F}_n] - \mathbb{E}[X|\mathcal{F}_n]| \leq \mathbb{E}[|X_n - X||\mathcal{F}_n] \to 0 \text{ a.s.} \qquad \text{as } n \to \infty$$

By the Upward Theorem once more, we see that $\mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}]$ a.s. as $n \to \infty$, and so

$$\mathbb{E}[X_n|\mathcal{F}_n] = (\mathbb{E}[X_n|\mathcal{F}_n] - \mathbb{E}[X|\mathcal{F}_n]) + \mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{G}] \text{ a.s.}$$

as $n \to \infty$.

⊣

2.6 Martingale Inequalities

The aim of this section is to state and prove two inequalities due to Doob.

Theorem 2.6.1 (Doob's Maximal Inequality)
Let X be a non–negative submartingale. Then for $c > 0$ in $\mathbb{R}$ and $n \in \mathbb{N}$, we have

$$c\mathbb{P}(\sup_{k\leq n} X_k \geq c) \leq \mathbb{E}[X_n; \sup_{k\leq n} X_k \geq c] \leq \mathbb{E}[X_n]$$

Proof: Let $F = \{\sup_{k\leq n} X_k \geq c\}$. Further inductively define a sequence $F_k, k \leq n$ by

$$\begin{aligned} F_0 &= \{X_0 \geq c\} \\ F_{k+1} &= \{X_0 < c\} \cap \{X_1 < c\} \cap \cdots \cap \{X_k < c\} \cap \{X_{k+1} \geq c\} \end{aligned}$$

Then F is the disjoint union $F = F_0 \cup \cdots \cup F_n$. Because X is adapted, we have $F_k \in \mathcal{F}_k$, and $X_k \geq c$ on F_k. Because X is a submartingale, we therefore have

$$c\mathbb{P}(F_k) \leq \mathbb{E}[X_k; F_k] \leq \mathbb{E}[X_n; F_k]$$

Summing over k, we obtain

$$c\mathbb{P}(F) \leq \mathbb{E}[X_n; F]$$

as required.

$\dashv$

Note that if M is a martingale, then M^2 is a non–negative submartingale. This follows easily from an application of Jensen's inequality. We thus have:

$$\mathbb{P}(\sup_{k\leq n} M_k \geq c) \leq \frac{1}{c^2}\mathbb{E}[M_n^2]$$

Before we prove the next theorem, recall Hölder's inequality: Suppose that $1 \leq p \leq \infty$, and that q is such that $\frac{1}{p} + \frac{1}{q} = 1$. If $X \in \mathcal{L}^p$ and $Y \in \mathcal{L}^q$, then $XY \in \mathcal{L}^1$, and

$$||XY||_1 \leq ||X||_p ||Y||_q$$

We need a lemma:

Lemma 2.6.2 *Suppose that X, Y are non–negative random variables such that*

$$c\mathbb{P}(X \geq c) \leq \mathbb{E}[Y; X \geq c] \quad \textit{for every } c > 0$$

If $p > 0$ and if $\frac{1}{p} + \frac{1}{q} = 1$, then

$$||X||_p \leq q||Y||_p$$

Proof: Define

$$\begin{aligned} I_1 &= \int_0^\infty pc^{p-1}\mathbb{P}(X \geq c)\, dc \\ I_2 &= \int_0^\infty pc^{p-2}\mathbb{E}[Y; X \geq c]\, dc \end{aligned}$$

Then clearly $I_1 \leq I_2$. Using Tonelli's Theorem[2], we change the order of integration:

$$\begin{aligned}
I_1 &= \int_0^\infty \left(\int_\Omega I_{\{X \geq c\}}(\omega)\, \mathbb{P}(d\omega) \right) pc^{p-1}\, dc \\
&= \int_\Omega \int_{c=0}^{X(\omega)} pc^{p-1}\, dc\, \mathbb{P}(d\omega) \\
&= \int_\Omega X^p(\omega)\, \mathbb{P}(d\omega) \\
&= \mathbb{E}[X^p]
\end{aligned}$$

Similarly,

$$\begin{aligned}
I_2 &= \int_0^\infty \left(\int_\Omega I_{\{X \geq c\}} Y\, \mathbb{P}(d\omega) \right) pc^{p-2}\, dc \\
&= \int_\Omega \left(\int_{c=0}^{X(\omega)} pc^{p-2}\, dc \right) Y\, \mathbb{P}(d\omega) \\
&= \mathbb{E}[\frac{p}{p-1} X^{p-1} Y] \\
&= \mathbb{E}[qX^{p-1}Y]
\end{aligned}$$

Now use Hölder's inequality to conclude that

$$\mathbb{E}[X^p] \leq \mathbb{E}[qX^{p-1}Y] \leq q||X^{p-1}||_q ||Y||_p$$

So far, we have not imposed any integrability conditions on X, Y. The lemma is obviously true if $||Y||_p = \infty$. Suppose now that $||X||_p, ||Y||_p < \infty$. Then since $(p-1)q = p$, we have

$$||X^{p-1}||_q = \mathbb{E}[X^p]^{\frac{1}{q}}$$

and thus

$$||X||_p \leq q||Y||_p$$

If $||Y||_p < \infty$, but $||X||_p = \infty$, replace X by $X \wedge n$ in the above. Note that the hypothesis of the lemma is still true, i.e.

$$c\mathbb{P}(X \geq c) \leq \mathbb{E}[Y; X \geq c] \Longrightarrow c\mathbb{P}(X \wedge n \geq c) \leq \mathbb{E}[Y; X \geq c]$$

and certainly $||X \wedge n||_p < \infty$. The result follows upon application of the Monotone Convergence Theorem as $n \to \infty$.

$\dashv$

Theorem 2.6.3 (Doob's $\mathcal{L}^p$–Inequality)
Suppose that $p > 1$ and that q is defined so that $\frac{1}{p} + \frac{1}{q} = 1$. Let X be a non–negative submartingale bounded in $\mathcal{L}^p$, and define

$$X^* = \sup_{k \in \mathbb{N}} X_k$$

Then $X^* \in \mathcal{L}^p$, and

$$||X^*||_p \leq q \sup_{k \in \mathbb{N}} ||X_k||_p$$

The submartingale X is therefore dominated by $X^* \in \mathcal{L}^p$

[2] i.e. Fubini's Theorem with non–negative integrands.

Proof: Define $X_n^* = \sup_{k \leq n} X_k$. Then $X_n^* \uparrow X^*$ a.s. By Doob's maximal inequality, we have $c\mathbb{P}(X_n^* \geq c) \leq \mathbb{E}[X_n; X_n^* \geq c]$ for every $c > 0$. By the preceding lemma, we therefore have $||X_n^*||_p \leq q||X_n||_p \leq q \sup_{k \in \mathbb{N}} ||X_k||_p$ for all n. Now apply the Monotone Convergence Theorem as $n \to \infty$ to obtain $||X^*||_p \leq q \sup_{k \in \mathbb{N}} ||X_k||_p$.

$\dashv$

We shall usually apply the $\mathcal{L}^p$–inequality for the case $p = 2$.

2.7 Continuous–Parameter Martingales

2.7.1 Stopping Times

Many of the concepts and results for discrete–parameter martingales can be extended to continuous–parameter martingales, and we shall spend some time doing so. *From now on, we shall assume that all martingales are at least cádlág, and that all filtrations satisfy the usual conditions.*

First we define the notion of a *stopping time* in an obvious way:

Definition 2.7.1 A random variable $\tau : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ is called a stopping time (w.r.t. a filtration $\mathcal{F}_t, t \geq 0$) if and only if

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \qquad \text{for each } t \geq 0$$

If X is a stochastic process, then we define the stopped variable X_τ by $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$. The σ–algebra of events prior to τ, denoted $\mathcal{F}_\tau$ and also called the *pre–τ algebra*, consist of all those events A with the property that

$$A \cap \{\tau \leq t\} \in \mathcal{F}_t \qquad \text{for all } t \geq 0$$

In the discrete framework, we saw that we could replace $\{\tau \leq t\} \in \mathcal{F}_t$ by $\{\tau = t\} \in \mathcal{F}_t$ in the definition of a stopping time. However, this does not work in the continuous case: If τ is a continuous random variable, then we clearly have $\mathbb{P}(\tau = t) = 0$, so $\{\tau = t\} \in \mathcal{F}_0 \subseteq \mathcal{F}_t$ always.

The result of the following exercise is often useful:

Exercise 2.7.2 One frequently also encounters the notion of an *optional time*. If $\mathcal{F}_t$ is a filtration which does not necessarily satisfy the usual conditions and if τ is a random variable, we say that τ is an optional time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$.

(i) Show that any stopping time is an optional time.

(ii) Show that if $\mathcal{F}_t$ satisfies the usual conditions, then any optional time is also a stopping time.

[Hint: (i) $\{\tau < t\} = \bigcup_n \{\tau \leq t - \frac{1}{n}\}$; (ii) $\{\tau \leq t\} = \bigcap_n \{\tau < t + \frac{1}{n}\} \in \mathcal{F}_s$ whenever $s > t$.]

$\square$

The next theorem lumps together all the basic results about stopping times:

Theorem 2.7.3 *Let σ, τ, τ_n be stopping times with respect to a filtration $\mathcal{F}_t$ satisfying the usual conditions.*

(a) $\sup_n \tau_n, \inf_n \tau_n, \limsup_n \tau_n, \liminf_n \tau_n, \lim_n \tau_n$ *are all stopping times.*

(b) $\tau \wedge \sigma, \tau \vee \sigma$ *and* $\tau + \sigma$ *are stopping times.*

(c) The pre-τ σ–algebra $\mathcal{F}_\tau$ is a σ–algebra which contains all the null sets, and τ is $\mathcal{F}_\tau$–measurable.

(d) If $\tau(\omega) = t$ for all ω, then $\mathcal{F}_\tau = \mathcal{F}_t$.

(e) If $\sigma \le \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

(f) A random variable X is $\mathcal{F}_\tau$–measurable if and only if $XI_{\{\tau \le t\}}$ is $\mathcal{F}_t$–measurable for each $t \ge 0$.

(g) If $A \in \mathcal{F}_\sigma$, then $A \cap \{\sigma \le \tau\}, A \cap \{\sigma < \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$.

(h) $\mathcal{F}_{\tau\wedge\sigma} = \mathcal{F}_\tau \cap \mathcal{F}_\sigma$.

(i) Each of the events $\{\tau < \sigma\}, \{\sigma < \tau\}, \{\tau \le \sigma\}, \{\sigma \le \tau\}, \{\tau = \sigma\}$ belongs to $\mathcal{F}_\tau \cap \mathcal{F}_\sigma$.

(j) If $\tau_n \downarrow \tau$, then $\mathcal{F}_{\tau_n} \downarrow \mathcal{F}_\tau$.

(k) If $\hat{\tau} : \Omega \to \infty$ is a map such that $\hat{\tau} = \tau$ a.s., then $\hat{\tau}$ is a stopping time, and $\mathcal{F}_{\hat{\tau}} = \mathcal{F}_\tau$.

Exercise 2.7.4 (1.) Prove the preceding Theorem.
[Hints:

(a) $\{\sup_n \tau_n \le t\} = \bigcap_n \{\tau_n \le t\}$; $\{\inf_n \tau_n < t\} = \bigcup_n \{\tau_n < t\}$, and use Exercise 2.1.11.

(b) $\tau \wedge \sigma = \inf\{\tau, \sigma\}$; $\{\tau + \sigma < t\} = \bigcup_{q,r \in \mathbb{Q}^+, q+r<t} \{\tau < q, \sigma < r\}$.

(c) $A^c \cap \{\tau \le t\} = (A \cap \{\tau \le t\})^c \cap \{\tau \le t\}$; If $N \in \mathcal{F}$ is a null set, then $N \cap \{\tau \le t\}$ is a null set, and thus in $\mathcal{F}_t$; $\{\tau \le s\} \cap \{\tau \le t\} = \{\tau \le s \wedge t\} \in \mathcal{F}_t$ for all s, t.

(d) For each s, $A \cap \{\tau \le s\}$ is either $\emptyset$ or A, depending on whether $s < t$ or $s \ge t$.

(e) If $\sigma \le \tau$, then $A \cap \{\tau \le t\} = A \cap \{\sigma \le t\} \cap \{\tau \le t\}$.

(f) $\Rightarrow$: Follow the usual procedure from indicator functions to simple to non–negative measurable functions etc.; $\Leftarrow$: Assume that $XI_{\{\tau \le t\}}$ is $\mathcal{F}_t$–measurable for all t. Note that $\{X \le r\} \cap \{\tau \le t\} = \{XI_{\{\tau \le t\}} \le r\} \cap \{\tau \le t\} \in \mathcal{F}_t$.

(g) Note that if τ is a stopping time, then $\tau \wedge t$ is $\mathcal{F}_t$–measurable, because $\{\tau \wedge t > s\} = \{\tau > s\} \cap \{t > s\}$. Next note that $[A \cap \{\sigma \le \tau\}] \cap \{\tau \le t\} = [A \cap \{\sigma \le t\}] \cap \{\tau \le t\} \cap \{\sigma \wedge t \le \tau \wedge t\}$, and use the just proven fact that both $\sigma \wedge t, \tau \wedge t$ are $\mathcal{F}_t$–measurable. The result with $<$ instead of $\le$ is proven similarly.

(h) If $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, then note that

$$A \cap \{\sigma \wedge \tau \le t\} = (A \cap \{\sigma \le \tau\} \cap \{\sigma \le t\}) \cup (A \cap \{\tau \le \sigma\} \cap \{\tau \le t\})$$

and use (g).

(i) By (g) and (h).

(j) Note that $A \cap \{\tau < t\} = \bigcup_n (A \cap \{\tau_n < t\})$.

(k) $\hat{\tau}$ is measurable, since $\hat{\tau}^{-1}(B)$ differs from $\tau^{-1}(B)$ by a null set.]

(2.) Prove that if $\tau_n \uparrow \tau$, then we do not necessarily have $\mathcal{F}_\tau = \sigma(\bigcup_n \mathcal{F}_{\tau_n})$.
[Hint: Let $X_t = 0$ for $0 \le t < 1$, let X_1 be a Bernoulli variable with $\mathbb{P}(X_1 = 1) = \frac{1}{2} = \mathbb{P}(X_1) = -1$, and let $X_t = X_1$ for $X_t > 1$. Let $\mathcal{F}_t$ be the natural augmented filtration, and let $\tau_n = 1 - \frac{1}{n}$.]

(3.) Show that if $\tau_n \uparrow \tau$ and $\bigcup_n \{\tau_n = \tau\} = \Omega$, then $\mathcal{F}_{\tau_n} \uparrow \mathcal{F}_\tau$.
[Hint: If $A \in \mathcal{F}_\tau$, then $A \cap \{\tau \le \tau_n\} = A \cap \{\tau = \tau_n\} \in \mathcal{F}_{\tau_n}$ by (g). Thus $A = \bigcup_n (A \cap \{\tau = \tau_n\}) \in \sigma\left(\bigcup_n \mathcal{F}_{\tau_n}\right)$.]

□

The next proposition will be very useful in transferring results from discrete–time to continuous–time. It states that any stopping time can be approximated (from above) by stopping times that take only countably many values. Recall that $[x]$ denoted the greatest integer less than or equal to x, i.e. $[x] = \sup \mathbb{Z} \cap (-\infty, x]$. Also define $[\infty] = \infty$.

Proposition 2.7.5 (Discretization Lemma) *Let τ be a stopping time. For each integer $n \ge 1$, define*

$$\tau_n(\omega) = \frac{[2^n \tau(\omega)] + 1}{2^n}$$

Then each τ_n is a stopping time, with $\tau_n \downarrow\downarrow \tau$ (pointwise).

Proof: Note that $\tau_n(\omega) = \frac{k+1}{2^n}$ whenever $k/2^n \le \tau(\omega) < (k+1)/2^n$, so that $\tau_n(\omega) > \tau(\omega)$ for each $\omega \in \Omega$. So define $\mathbb{D}_n^+ = \{k/2^n : k = 0, 1, 2, \dots\}$ (the set of non–negative dyadic rationals of order $\le n$), and for each n, define maps $a_n, b_n : \mathbb{R}^+ \to \mathbb{D}_n^+$ by

$$a_n(t) = \max\{d \in \mathbb{D}_n^+ : d \le t\} \qquad b_n(t) = \min\{d \in \mathbb{D}_n^+ : d > t\}$$

so that $a_n(t) \le t < b_n(t)$ for each n. Since $\mathbb{D}_n^+ \subseteq \mathbb{D}_{n+1}^+$, we see that $a_n(t) \uparrow t$ and $b_n(t) \downarrow\downarrow t$ (as $n \to \infty$). Moreover, $\tau_n(\omega) = b_n(\tau(\omega))$, by definition, and each $\mathbb{D}_n^+$ is countable. Hence the range of τ_n is countable. Now note that $\tau_n(\omega) \le t$ if and only if $\tau(\omega) < a_n(t)$, so that

$$\{\tau_n \le t\} = \{\tau < a_n(t)\} \in \mathcal{F}_{a_n(t)} \subseteq \mathcal{F}_t$$

This proves that each τ_n is a stopping time.

⊣

Exercise 2.7.6 If we try to approximate τ by stopping times $\tau_n \uparrow \tau$ from below by putting $\tau_n(\omega) = a_n(\tau(\omega))$ (where a_n is defined as in the proof of the Discretization Lemma), then we run into trouble. Why?

□

In the discrete context, we proved in Exercise 2.2.8 that if τ is a stopping time and X an adapted stochastic process, then X_τ is $\mathcal{F}_\tau$–measurable. However, this proof depends on the random variable X_τ taking only countably many values. For example, if τ is integer–valued, then using $X_\tau^{-1}(B) = \bigcup_n X_n^{-1}(B) \cap \{\tau = n\}$, we see that $X_\tau^{-1}(B) \in \mathcal{F}_\tau$ for any Borel set B. This argument will obviously not work in continuous–time. However, assuming that the process X_t is *right–continuous*, we can use the Discretization Lemma to prove that X_τ *is* $\mathcal{F}_\tau$–measurable in the continuous–time case as well:
Choose stopping times $\tau_n \downarrow \tau$ as in the Discretization Lemma, so that each τ_n has a countable range. It is then easy to prove that each X_{τ_n} is $\mathcal{F}_{\tau_n}$–measurable, as in the discrete–time case. By Theorem 2.7.3(j), $\mathcal{F}_{\tau_n} \downarrow \mathcal{F}_\tau$. Since $\limsup_n X_{\tau_n} = \limsup_{n \geq N} X_{\tau_n}$, it follows that $\limsup_n X_{\tau_n}$ is $\mathcal{F}_{\tau_N}$–measurable for each N, and thus that it is $\mathcal{F}_\tau$–measurable. But since X_t is right–continuous a.s., we see that $\limsup_n X_{\tau_n} = X_\tau$ a.s. Thus, using the fact that $\mathcal{F}_\tau$ contains all the $\mathbb{P}$–null sets, it follows that X_τ is $\mathcal{F}_\tau$–measurable.

We have thus shown:

Proposition 2.7.7 *Let $(X_t)_{0 \leq t \leq \infty}$ be an adapted process with a.s. right–continuous sample paths. If τ is a stopping time, then X_τ is $\mathcal{F}_\tau$–measurable.*

2.7.2 Martingales in Continuous Time

In this section, we will generalize the main discrete–parameter martingale results from Chapter 1 to the continuous–parameter cádlág case. Throughout, all filtrations are assumed to satisfy the usual conditions. Our first aim is to prove that $\mathcal{L}^1$–bounded sub/supermartingales converge. The fact that $\mathbb{Q}^+$ is dense in $\mathbb{R}^+$ will be very important. Typically, $\mathbb{Q}$ will be written as a countable union of an increasing family of finite sets. Restricted to each of these finite sets, the stochastic processes look like discrete–parameter processes, and all the discrete–parameter results will hold. Creative use of the monotonicity properties of measure and the integration theorems will allow us to extend the results to stochastic processes indexed by $\mathbb{Q}^+$. Finally, right–continuity will be used to extend the results to stochastic processes indexed by $\mathbb{R}^+$. Since right–continuity involves approximation from above, the Lévy–Doob results on reversed martingales will be important.

We begin by defining the notion of the *number of upcrossings*, which, as in the discrete–parameter case, will come in very handy.

Definition 2.7.8 (a) Suppose that $X_t, t \geq 0$ is a real–valued adapted stochastic process, and let F be a finite subset of $\mathbb{R}^+$ For $a < b \in \mathbb{R}$, define $U_F(X; [a, b])$ to be the number of upcrossings of $[a, b]$ in F. To be precise, define a double sequence of stopping times τ_k, σ_k recursively by

$$\begin{aligned} \tau_1(\omega) &= \min\{t \in F : X_t(\omega) < a\} \\ \sigma_j(\omega) &= \min\{t \in F : t \geq \tau_j(\omega), X_t(\omega) > b\} \\ \tau_{j+1}(\omega) &= \min\{t \in F : t \geq \sigma_j(\omega), X_t(\omega) < a\} \end{aligned}$$

We use here the convention that $\min(\emptyset) = +\infty$. Then $U_F(X; [a, b])(\omega)$ is defined to be the largest integer n for which $\sigma_n(\omega) < \infty$.

(b) If $I \subseteq \mathbb{R}^+$ is not necessarily finite, we define

$$U_I(X; [a, b]) = \sup\{U_F(X; [a, b]) : F \text{ a finite } \subseteq I\}$$

As in the discrete–parameter case, we have:

Theorem 2.7.9 (Upcrossing Lemma)
Suppose that X is a cádlág supermartingale, let $[S, T]$ be a subinterval of $\mathbb{R}^+$, and let $a < b \in \mathbb{R}$. Then

$$\mathbb{E}U_{[S,T]}(X; [a, b]) \leq \frac{\mathbb{E}[(X_T - a)^-]}{b - a}$$

Proof: We can prove this directly from the discrete–parameter version of the Upcrossing Lemma. Let F_n be an increasing sequence of finite subsets of $[S, T]$ with the following properties:

(a) $S, T \in F_n$ for all n.

(b) $\bigcup_n F_n = [S, T] \cap \mathbb{Q}$

Now note that since x is cádlág, we must have $U_{[S,T]} = U_{[S,T]\cap\mathbb{Q}} =\uparrow \lim_n U_{F_n}$, and thus that $\mathbb{E}_{[S,T]} = \lim_n \mathbb{E}U_{F_n}$, by the Monotone Convergence Theorem. But $(X_t : t \in F_n)$ is a discrete–parameter supermartingale (for each n), and $T \in F_n$. Thus by the discrete version of the Upcrossing Lemma, we have $(b - a)\mathbb{E}U_{F_n} \leq \mathbb{E}[(X_T - a)^-]$. Taking limits yields the result.

$\dashv$

We now prove the respective martingale convergence theorems in one fell swoop. Recall that the notion of *uniform integrability* was defined for arbitrary collections of random variables, and that we do not need to redefine it for continuous–parameter processes. The same goes for the notion of *uniform $\mathbb{P}$–continuity.*

Theorem 2.7.10 (Doob's Martingale Convergence Theorem)

(a) Let X be a cádlág supermartingale bounded in $\mathcal{L}^1$. Then there is a random variable X_∞ such that

$$X_t \to X_\infty \quad \text{a.s.}$$

and $\mathbb{E}|X_\infty| < \infty$.

(b) Moreover, if X is UI*, then $X_t \to X_\infty$ in $\mathcal{L}^1$, and then $\mathbb{E}[X_\infty|\mathcal{F}_t] \leq X_t$.*

(c) Finally, if X is a martingale, then $X_t \to X_\infty$ in $\mathcal{L}^1$ if and only if X_t is UI*, and then $X_t = \mathbb{E}[X_\infty|\mathcal{F}_t]$ in (b).*

Proof: (a) Define $C = \sup_t \mathbb{E}|X_t|$, so that $C < \infty$ (by assumption). Further define:

$$X_\infty^+(\omega) = \limsup_{t\to\infty} X_t(\omega) \qquad X_\infty^-(\omega) = \liminf_{t\to\infty} X_t(\omega)$$

Now if $\lim_{t\to\infty} X_t(\omega)$ does not exist, then we can find real numbers a, b such that $X_\infty^-(\omega) < a < b < X_\infty^+(\omega)$, and thus $U_{\mathbb{R}^+}(X(\omega);[a,b]) = \infty$. However,

$$\mathbb{E}U_{[0,n]}(X(\omega);[a,b]) \leq \frac{\mathbb{E}[(X_n - a)^-]}{b-a} \leq \frac{C + |a|}{b-a} < \infty$$

Letting $n \to \infty$, we see that

$$\mathbb{E}U_{\mathbb{R}^+}(X;[a,b]) < \frac{C+|a|}{b-a} < \infty$$

by the Monotone Convergence Theorem. It follows that the set $\{\omega : U_{\mathbb{R}^+}(X(\omega);[a,b]) = \infty\}$ has measure zero, i.e. $X_\infty(\omega) = \lim_{t\to\infty} X_t(\omega)$ exists a.s. Moreover, by Fatou's Lemma, $\mathbb{E}|X_\infty| \leq \liminf_t \mathbb{E}|X_t| < C < \infty$.

To prove (b), note that if X is UI and convergent in probability, then it is $\mathcal{L}^1$–convergent, just as in the discrete case.[3] Since we also have $X_t = \mathbb{E}[X_n|\mathcal{F}_t]$ whenever $n \geq t$, and since $X_n \geq \mathbb{E}[X_\infty|\mathcal{F}_n]$ by the discrete–parameter result, we see upon application of a conditional expectation with respect to $\mathcal{F}_t$ that

$$X_t \geq \mathbb{E}[X_\infty|\mathcal{F}_t]$$

as required.

Finally, we can prove (c) as follows: By (b) we know that if X is UI, then $X_t \to X_\infty$ in $\mathcal{L}^1$, because every martingale is a supermartingale, and the same argument will prove that $X_t = \mathbb{E}[X_\infty|\mathcal{F}_t]$. Suppose now that $X_t \to X_\infty$ in $\mathcal{L}^1$. Then clearly also $\mathbb{E}[X_t|\mathcal{G}] \to \mathbb{E}[X_\infty|\mathcal{G}]$ in $\mathcal{L}^1$, for every $\mathcal{G} \subseteq \mathcal{F}$. It follows that $X_s = \mathbb{E}[X_t|\mathcal{F}_s] \to \mathbb{E}[X_\infty|\mathcal{F}_s]$ in $\mathcal{L}^1$ as $t \to \infty$, and thus that each $X_s = \mathbb{E}[X_\infty|\mathcal{F}_s]$. But then X_t is UI, by Theorem 2.4.8.

$\dashv$

[3] As an exercise, check that the proof (c) $\Rightarrow$ (d) $\Rightarrow$ (a) in Theorem 2.4.5 also holds for continuous–parameter processes. However, if you peruse the proof of (b) $\Rightarrow$ (c), you will note that a discrete–parameter process is essential here. In fact, it is not true in general that an $\mathcal{L}^1$-convergent supermartingale X is UI in continuous–time, although this *is* the case if X is a martingale.

It follows that a continuous–parameter martingale M_t is UI precisely if it is of the form $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$ for some integrable random variable M_∞. All UI martingales are of this form.

Most of our results deal with right–continuous (super–, sub–)martingales. However, results such as the Upcrossing Lemma do actually imply some fairly strong continuity conditions. We shall show that if X_t is a submartingale (w.r.t. some filtration satisfying the usual conditions), then X_t has a right–continuous version if and only if the map $t \mapsto \mathbb{E}X_t$ is right—continuous. First, we need a lemma:

Lemma 2.7.11 *Let X_t be an $\mathcal{F}_t$–submartingale. (We do not assume that X_t is right–continuous, nor that $\mathcal{F}_t$ satisfies the usual conditions.)*

(i) There is an event $\Omega^* \in \mathcal{F}$ of measure 1 such that for every $\omega \in \Omega^*$ the limits

$$X_{t^+}(\omega) = \lim_{q \downarrow t, q \in \mathbb{Q}} X_q(\omega) \qquad X_{t^-}(\omega) = \lim_{q \uparrow t, q \in \mathbb{Q}} X_q(\omega)$$

exist for all $t \geq 0$ (resp. $t > 0$).

(ii) Moreover,

$$\begin{aligned} \mathbb{E}[X_{t^+}|\mathcal{F}_t] &\geq X_t \qquad \text{a.s.} \\ \mathbb{E}[X_t|\mathcal{F}_{t^-}] &\geq X_{t^-} \qquad \text{a.s.} \end{aligned}$$

for all $t \geq 0$ (resp. $t > 0$).

(iii) X_{t^+} is a $\mathcal{F}_{t^+}$–submartingale with almost every sample path cádlág.

(The set $\mathbb{Q}$ may be replaced by any countable dense subset of $\mathbb{R}$.)

Proof: (i) For $a < b \in \mathbb{Q}$ and $n \in \mathbb{N}$, define

$$A^{(n)}_{a,b} = \{\omega \in \Omega : U_{[0,n]\cap\mathbb{Q}}(X(\omega); [a,b]) = \infty\},$$

Arguing as in the proof of Upcrossing Lemma, we see that $\mathbb{P}(A^{(n)}_{a,b}) = 0$, for otherwise $\mathbb{E}U_{[0,n]\cap\mathbb{Q}}(X(\omega); [a,b]) = \infty$.
Now if $t \geq 0$, choose $n \in \mathbb{N}$ such that $t \leq n$. If $a = \liminf_{q\downarrow t, q\in\mathbb{Q}} X_q(\omega) < b = \limsup_{q\downarrow t, q\in\mathbb{Q}} X_q(\omega)$, then $\omega \in A^{(n)}_{a,b}$. Thus $\lim_{q\downarrow t, q\in\mathbb{Q}} X_q(\omega)$ exists for almost all ω. Similarly, $\lim_{q\uparrow t, q\in\mathbb{Q}} X_q(\omega)$ exists for almost all ω.

(ii) Let $q_n \downarrow t$ (strictly) where $q_n \in \mathbb{Q}$. Put $Y_n = X_{q_n}$, $\mathcal{G}_n = \mathcal{F}_{q_n}$. Then Y_n is a reversed $\mathcal{G}_n$–submartingale, and $Y_n \to X_{t^+}$ a.s. (by definition of X_{t^+} in (i)). Moreover, $\mathbb{E}Y_n \geq \mathbb{E}X_t > -\infty$ for each n. By the Lévy–Doob Downward Theorem, the family Y_n is UI. Since $Y_n \to X_t$ a.s., we now have $Y_n \to X_t$ in $\mathcal{L}^1$ as well, by Theorem 2.4.5.

Now if $A \in \mathcal{F}_t$, then $\mathbb{E}[X_t; A] \leq \mathbb{E}[Y_n; A]$, by the submartingale property, so that $\mathbb{E}[X_t; A] \leq \mathbb{E}[X_{t^+}; A]$, since $Y_n \to X_t$ in $\mathcal{L}^1$. Thus we have shown that $\mathbb{E}[X_{t^+}|\mathcal{F}_t] \geq X_t$.

To prove the second inequality, choose $r_n \in \mathbb{Q}$ with $r_n \uparrow t$ (strictly). Then by the Strong Dominated Convergence Theorem (Theorem 2.5.5), $\mathbb{E}[X_t|\mathcal{F}_{r_n}] \to \mathbb{E}[X_t|\mathcal{F}_{t^-}]$ as $n \to \infty$, and by the submartingale property, $X_{r_n} \leq \mathbb{E}[X_t|\mathcal{F}_{r_n}]$. Letting $n \to \infty$, we get the second inequality.

(iii) it is easy to see that X_{t^+} is adapted to $\mathcal{F}_{t^+}$. Moreover, if $s < t$ and $q_n \downarrow s$ (with each $q_n < t$), then $\mathbb{E}[X_{t^+}|\mathcal{F}_{q_n}] \geq \mathbb{E}[X_t|\mathcal{F}_{q_n}] \geq X_{q_n}$ a.s., by the first inequality of (ii) and the

submartingale property. By the Strong Dominated Convergence Theorem, we therefore have

$$\mathbb{E}[X_{t^+}|\mathcal{F}_{t^+}] = \lim_n \mathbb{E}[X_{t^+}|\mathcal{F}_{q_n}] \geq \lim_n X_{q_n} = X_{s^+} \quad \text{a.s.}$$

Thus X_{t^+} is indeed an $\mathcal{F}_{t^+}$–submartingale. To see that almost every sample path of X_{t^+} is cádlág is now easy, using (i).

⊣

Having proved the lemma, we can now give conditions under which a submartingale has a cádlág modification:

Theorem 2.7.12 *Let X_t be a submartingale w.r.t. a filtration $\mathcal{F}_t$ satisfying the usual conditions. Then X_t has a cádlág modification (which is also a submartingale adapted to $\mathcal{F}_t$) if and only if the function $t \mapsto \mathbb{E}X_t$ is right–continuous.*

Proof: First assume that the map $t \mapsto \mathbb{E}X_t$ is indeed right–continuous. Since $\mathcal{F}_{t^+} = \mathcal{F}_t$ (by the usual conditions), the process X_{t^+} defined in Lemma 2.7.11 is a $\mathcal{F}_t$–submartingale with a.s. cádlág sample paths. It therefore remains to show that X_{t^+} is a version of X_t, i.e. that $\mathbb{P}(X_t = X_{t^+}) = 1$ for all $t \geq 0$. So let $t \geq 0$ and choose rational numbers $q_n \downarrow t$. By the Lévy–Doob Downward Theorem, the reversed submartingale $(X_{q_n}, \mathcal{F}_{q_n})$ is UI. Since $X_{q_n} \to X_{t^+}$ a.s., we therefore must have $\lim_n \mathbb{E}X_{q_n} = \mathbb{E}X_{t^+}$ by Theorem 2.4.5. But by hypothesis, $\mathbb{E}X_t = \lim_n \mathbb{E}X_{q_n}$, and so $\mathbb{E}X_{t^+} = \mathbb{E}X_t$. However, by Lemma 2.7.11 (ii), $X_{t^+} \geq X_t$ a.s. (since $\mathcal{F}_{t^+} = \mathcal{F}_t$). Thus $X_{t^+} = X_t$ a.s.

For the converse, suppose that if Y_t is a right–continuous modification of X_t. Pick $t \geq 0$, and let $t_n \downarrow t$. We must show that $\mathbb{E}X_t = \lim_n \mathbb{E}X_{t_n}$. Now because Y_t is a modification of X_t, we have

$$\mathbb{P}(X_t = Y_t, \forall n \geq 0 (X_{t_n} = Y_{t_n})) = 1$$

(since this involves a countable intersection of sets of measure 1). Also $Y_t = \lim_n Y_{t_n}$ a.s., by right–continuity of Y_t. It follows that $X_t = \lim_n X_{t_n}$ a.s. But, again by the Lévy–Doob Downward Theorem, the family X_{t_n} is UI. Hence $\mathbb{E}X_t = \lim_n \mathbb{E}X_{t_n}$ by Theorem 2.4.5.

⊣

Immediately, we have:

Corollary 2.7.13 *Let X_t be a martingale with respect to a filtration that satisfies the usual conditions. Then X_t has a cádlág modification.*

Next, we tackle continuous–parameter versions of the martingale inequalities.

Theorem 2.7.14 (Doob's Maximal Inequality)
Let X_t be a cádlág submartingale, let $[S, T]$ be a subinterval of $\mathbb{R}^+$, and let $c > 0$ be a real numbers. Then

$$c\mathbb{P}(\sup_{S \leq t \leq T} X_t \geq c) \leq \mathbb{E}[X_T^+]$$

Proof: Note that if X is a submartingale, then X^+ is a non–negative submartingale. This follows from the fact that $\varphi(x) = x^+ = \max\{x, 0\}$ is convex, and Jensen's inequality. Let F'

be a finite subset of $[S,T]\cap\mathbb{Q}$, and let $F = F' \cup \{S,T\}$. Also let $0 < c' < c$. By the discrete version of Doob's maximal inequality, we have

$$c'\mathbb{P}(\sup_{t\in F} X_t > c') \le c'\mathbb{P}(\sup_{t\in F} X_t^+ > c') \le \mathbb{E}X_T^+$$

This is true for all finite $F' \subseteq [S,T]\cap\mathbb{Q}$. Now choose an increasing sequence F_n' of finite subsets of $\mathbb{Q}$ with the property that $\bigcup_n F_n' = [S,T]\cap\mathbb{Q}$, and let $F_n' = F_n \cup \{S,T\}$ for each n. Let $G = \bigcup_n F_n$. We now see that

$$c'\mathbb{P}(\sup_{t\in G} X_t > c') \le \mathbb{E}[X_T^+]$$

as follows: Note that

$$\{\sup_{t\in G} X_t > c'\} = \bigcup_n A_n$$

where

$$A_n = \{\sup_{t\in F_n} X_t > c'\}$$

Note that A_n is increasing, and thus it follows by monotonicity properties of measure that $\mathbb{P}(\{\sup_{t\in G} X_t > c'\}) = \lim_n \mathbb{P}(A_n)$. But $c'\mathbb{P}(A_n) \le \mathbb{E}X_T^+$, and thus $c'\mathbb{P}(\{\sup_{t\in G} X_t > c'\}) \le \mathbb{E}X_T^+$. By right–continuity, we have $c'\mathbb{P}(\{\sup_{S\le t\le T} X_t > c'\}) \le \mathbb{E}X_T^+$. Finally, letting $c' \uparrow c$ yields the result.

⊣

Theorem 2.7.15 (Doob's $\mathcal{L}^p$–inequality) *Suppose that $1 < p < \infty$ and that q is defined so that $\frac{1}{p} + \frac{1}{q} = 1$. Let X be a non–negative submartingale bounded in $\mathcal{L}^p$, and define*

$$X^* = \sup_{t\ge 0} X_t$$

Then $X^* \in \mathcal{L}^p$, and

$$||X^*||_p \le q \sup_{t\ge 0} ||X_t||_p$$

The submartingale X is therefore dominated by $X^* \in \mathcal{L}^p$.

Proof: The proof of the discrete–parameter version depends on Doob's maximal inequality, which we have just established. Thus the proof of the discrete–parameter theorem will also work for the continuous–parameter version.

⊣

Doob's $\mathcal{L}^p$–inequality is very useful for proving convergence results. For example, suppose that X_t is a martingale bounded in $\mathcal{L}^p$, where $p > 1$. By Jensen's inequality, $|X_t|^p$ is a non–negative submartingale, and by the $\mathcal{L}^p$–inequality, $|X_t|^p$ is dominated by the integrable random variable $(X^*)^p$. Thus if $X_t \to X_\infty$ a.s., then $|X_t - X_\infty|^p \to 0$ a.s., and $|X_t - X_\infty|^p$ is dominated by $(2X^*)^p \in \mathcal{L}^1$. By the Dominated Convergence Theorem, we thus have $X_t \to X_\infty$ in $\mathcal{L}^p$ as well.

The following corollary is just a reformulation of Doob's $\mathcal{L}^p$–inequality:

Corollary 2.7.16 *Suppose that $1 < p < \infty$ and that M is a martingale bounded in $\mathcal{L}^p$. then*

$$\mathbb{E}[\sup_{0 \le t \le T} |M_t|^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|M_T|^p]$$

In particular,

$$\mathbb{E}[\sup_{0 \le t \le T} M_t^2] \le 4\mathbb{E}[M_T^2].$$

$\dashv$

2.7.3 Optional Sampling

We now turn our attention to a continuous–parameter version of the optional sampling theorem. In order to be able to use the discrete–parameter version, we need to be able to approximate a cádlág martingale by a discrete version. As in the proof of the Discretization Lemma, we define, for each $n \in \mathbb{N}$,

$$\mathbb{D}_n^+ = \{k2^{-n} : k \in \mathbb{N}\}$$

to be the set of *non–negative dyadic rationals of order* $\le n$. Note that $m \le n$ implies $\mathbb{D}_m^+ \subseteq \mathbb{D}_n^+$, and that $\bigcup_n \mathbb{D}_n^+$ is a countable dense subset of $\mathbb{R}^+$. Note also that if X is a martingale, then $(X_q : q \in \mathbb{D}_n^+)$ is a discrete–parameter martingale, for each n.

Lemma 2.7.17 *Let X be a cádlág supermartingale, and let τ be a stopping time. Let $t \ge 0$, and define*

$$\tau_n(\omega) = \inf\{q \in \mathbb{D}_n^+ : q > \tau(\omega)\} \qquad t_n = \inf\{q \in \mathbb{D}_n^+ : q > t\}$$

Then τ_n is a stopping time w.r.t. the filtration $(\mathcal{F}_q : q \in \mathbb{D}_n^+)$. We also have $\tau_n \downarrow \tau$ and $\mathcal{F}_{\tau_n} \downarrow \mathcal{F}_\tau$. Moreover

$$X_{\tau_n \wedge t_n} \to X_{\tau \wedge t} \quad \text{a.s. } \textit{and in } \mathcal{L}^1$$

and thus $X_{\tau \wedge t} \in \mathcal{L}^1$.

Proof: The τ_n defined here is exactly the same as the τ_n defined in the Discretization Lemma. Thus each τ_n is a stopping time and $\tau_n \downarrow \tau$. It follows from Theorem 2.7.3 that $\mathcal{F}_{\tau_n} \downarrow \mathcal{F}_\tau$ as well. Note also that $t_n \downarrow t$, and thus that $\tau_n \wedge t_n \downarrow \tau \wedge t$.

Using the discrete version of the Optional Sampling Theorem (applied to the closed discrete–parameter supermartingale $(X_{k/2^{n+1}} : k/2^{n+1} \le t+1)$), we see that

$$\mathbb{E}[X_{\tau_n \wedge t_n} | \mathcal{F}_{\tau_{n+1} \wedge t_{n+1}}] \le X_{\tau_{n+1} \wedge t_{n+1}} \qquad \text{a.s.}$$

and similarly that $\mathbb{E}X_{\tau_n \wedge t_n} \le \mathbb{E}X_0$. So define $Y_n = X_{\tau_n \wedge t_n}$, $\mathcal{G}_n = \mathcal{F}_{\tau_n \wedge t_n}$. Then we have

$$\mathcal{G}_1 \supseteq \mathcal{G}_2 \supseteq \cdots \supseteq \mathcal{G}_n \supseteq \dots$$

and Y_n is a reversed $\mathcal{G}_n$–supermartingale. Moreover, $\lim_n \mathbb{E}Y_n \le \mathbb{E}X_0 < \infty$. By the Lévy–Doob Downward Theorem, Y is UI and the limit

$$X_{\tau \wedge t} = Y_\infty = \lim_{n \to \infty} Y_n$$

exists a.s. and in $\mathcal{L}^1$ (where we used the fact that X is cádlág).

⊣

Suppose that X is a continuous–parameter stochastic process, and that τ is a stopping time. We define the stopped process X^τ by $X^\tau_t = X_{\tau\wedge t}$. The following result cannot be unexpected:

Theorem 2.7.18 (Stopped (super)martingales are (super)martingales)
Suppose that X_t is a cádlág (super)martingale (w.r.t. a filtration that satisfies the usual conditions). If τ is a stopping time, then X^τ is also a cádlág (super)martingale w.r.t. the same filtration.

Proof: Suppose that X_t is a cádlág supermartingale, and let $0 \leq s \leq t$. Define τ_n, t_n as in the above lemma, and define s_n analogously (i.e. $s_n = \inf\{q \in \mathbb{D}^+_n : q > s\}$). By the discrete–parameter "Stopped (super)martingales are (super)martingales" theorem, it follows that

$$\mathbb{E}[X_{\tau_n\wedge t_n}|\mathcal{F}_{s_m}] \leq X_{\tau_n\wedge s_m} \qquad \text{for } m \geq n$$

(using the parameter set $\mathbb{D}^+_m$.) Now let $m \to \infty$. Note that $\lim_n \mathbb{E}X_{\tau_n\wedge t_n} \leq \mathbb{E}X_0 < \infty$, so by the Lévy–Doob Downward Theorem and right–continuity,

$$\mathbb{E}[X_{\tau_n\wedge t_n}|\mathcal{F}_s] \leq X_{\tau_n\wedge s}$$

Now let $n \to \infty$. Then

$$\mathbb{E}[X_{\tau_n\wedge t_n}|\mathcal{F}_s] \to \mathbb{E}[X_{\tau\wedge t}|\mathcal{F}_s]$$

by the preceding lemma, and

$$X_{\tau_n\wedge s} \to X_{\tau\wedge s}$$

by right–continuity. Putting the results together yields

$$\mathbb{E}[X_{\tau\wedge t}|\mathcal{F}_s] = X_{\tau\wedge s} \qquad \text{a.s.}$$

as required.

⊣

In the discrete–parameter theory, the preceding result was proved using a martingale transform. We have not yet defined continuous–parameter analogues of martingale transforms and previsible processes, but that is because these are difficult to generalize, and will need quite a bit more theory: The generalization of the martingale transform to continuous–time is the stochastic integral.

We can now prove:

Theorem 2.7.19 (Doob's Optional Sampling Theorem for closed cádlág supermartingales)
Suppose that X is a closed cádlág supermartingale, and let σ, τ be stopping times. Then $X_\tau, X_\sigma \in \mathcal{L}^1$, and

$$\mathbb{E}[X_\infty|\mathcal{F}_\tau] \leq X_\tau \quad \text{a.s.} \qquad \mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_{\tau\wedge\sigma} \quad \text{a.s.}$$

with equality if X is a closed martingale.

Proof: First assume that $\sigma \leq \tau$ a.s. Let X_∞ be a last element of X, i.e. X_∞ is $\mathcal{F}_\infty$–measurable and integrable, with $\mathbb{E}[X_\infty|\mathcal{F}_t] \leq X_t$ for all t (and equality if X is a martingale). Define τ_n, σ_n as in Lemma 2.2.10. Note that $\sigma_n \leq \tau_n$. Note that $\tau_n \downarrow \tau$, and thus that $X_{\tau_n} \to X_\tau$ a.s. (by right–continuity). Furthermore, $\mathcal{F}_{\tau_n} \downarrow \mathcal{F}_\tau$, by Theorem 2.7.3, and $\mathbb{E}X_{\tau_n} = \mathbb{E}X_{\tau_n \wedge \infty} \leq \mathbb{E}X_0 < \infty$ (by the discrete Optional Sampling Theorem applied to the stopped supermartingale $X_t^{\tau_n}$). Thus by the Lévy–Doob Downward Theorem, the family X_{τ_n} is UI, and thus $X_{\tau_n} \to X_\tau$ in $\mathcal{L}^1$ as well, and $\mathbb{E}[X_{\tau_n}|\mathcal{F}_\tau] \leq X_\tau$ a.s. (with equality if X is a martingale). It follows that X_τ is integrable. Similarly, the family X_{σ_n} is UI, and $X_{\sigma_n} \to X_\sigma$ a.s. and in $\mathcal{L}^1$, $\mathbb{E}[X_{\sigma_n}|\mathcal{F}_\sigma] \leq X_\sigma$ a.s., and X_σ is integrable.

By the discrete Optional Sampling Theorem (on the parameter set $\mathbb{D}_n^+$), we have

$$\mathbb{E}[X_{\tau_n}|\mathcal{F}_{\sigma_n}] \leq X_{\sigma_n} \quad \text{a.s.}$$

and by Proposition 2.5.4(b), $\mathbb{E}[X_{\tau_n}|\mathcal{F}_{\sigma_n}] \to \mathbb{E}[X_\tau|\mathcal{F}_\sigma]$ in $\mathcal{L}^1$. Hence there is a subsequence n_k such that $\mathbb{E}[X_{\tau_{n_k}}|\mathcal{F}_{\sigma_{n_k}}] \to \mathbb{E}[X_\tau|\mathcal{F}_\sigma]$ a.s. Since also $\mathbb{E}[X_{\tau_{n_k}}|\mathcal{F}_{\sigma_{n_k}}] \leq X_{\sigma_{n_k}}$ a.s., we obtain $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_\sigma$ a.s. upon letting $k \to \infty$ on both sides. We have now proved the Optional Sampling Theorem in case $\sigma \leq \tau$ a.s.

Now Let σ, τ be arbitrary. We want to show that $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leq X_{\tau\wedge\sigma}$. Since $\sigma \wedge \tau \leq \tau$, we can apply the version of the Optional Sampling Theorem just proved to deduce that $\mathbb{E}[X_\tau|\mathcal{F}_{\sigma\wedge\tau}] \leq X_{\sigma\wedge\tau}$. Now if $A \in \mathcal{F}_\sigma$, then $A \cap \{\sigma \leq \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$, by Theorem 2.7.3(g). Thus

$$\mathbb{E}[X_\tau; A] = \mathbb{E}[X_\tau; A \cap \{\sigma \leq \tau\}] + \mathbb{E}[X_\tau; A \cap \{\sigma > \tau\}]$$

But $\mathbb{E}[X_\tau; A \cap \{\sigma \leq \tau\}] \leq \mathbb{E}[X_{\tau\wedge\sigma}; A \cap \{\sigma \leq \tau\}]$ (with equality if X is a UI martingale) because $\mathbb{E}[X_\tau|\mathcal{F}_{\sigma\wedge\tau}] \leq X_{\sigma\wedge\tau}$ and $A \cap \{\sigma \leq \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$. Also, $X_\tau = X_{\tau\wedge\sigma}$ on the set $A \cap \{\sigma > \tau\}$. The result now follows.

⊣

Remarks 2.7.20 The following facts maybe useful when applying the Optional Sampling Theorem:

1. A UI supermartingale is closed (but not necessarily vice versa in the continuous–parameter case).

2. A martingale is closed iff it is UI.

□

Exercise 2.7.21 (a) If M is a martingale and K a finite non–negative real, then the stopped martingale M^K is UI.

(b) Let M be a martingale. Suppose that σ, τ are stopping times, and that the stopped martingale M_t^τ is UI. Then the stopped martingale $M_t^{\tau\wedge\sigma}$ is UI.

[Hint: (b) Note first that $M_t^\tau \to M_\tau$ a.s. and in $\mathcal{L}^1$ as $t \to \infty$ (why?) and thus that $M_\tau \in \mathcal{L}^1$. Next, note that $M_t^{\tau\wedge\sigma} = \mathbb{E}[M_\tau^\tau|\mathcal{F}_{\sigma\wedge t}] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge t}]$. Theorem 2.4.8]

□

Next, we characterize all UI cádlág martingales (cf. Theorem 2.4.14):

Theorem 2.7.22 *Suppose that M is an adapted cádlág process with $M_0 = 0$ such that for every stopping time $\tau \leq \infty$ we have*

$$\mathbb{E}|M_\tau| < \infty \qquad \mathbb{E}M_\tau = 0$$

Then M is a UI *martingale.*

Proof: With $\tau = \infty$, we see that $M_\infty \in \mathcal{L}^1$ (where $M_\infty(\omega)$ is *defined* to be $\lim_{t\to\infty} M_t(\omega)$ if this limit exists, and 0 otherwise). To show that M is a UI martingale, it suffices to check that

$$M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$$

So let $F \in \mathcal{F}_t$. Define

$$\tau(\omega) = \begin{cases} t & \text{if } \omega \in F \\ \infty & \text{else} \end{cases}$$

Then τ is easily seen to be a stopping time. Now

$$\mathbb{E}[M_\infty] = \mathbb{E}[M_\infty; F] + \mathbb{E}[M_\infty; \Omega - F] = 0$$
$$\mathbb{E}[M_\tau] = \mathbb{E}[M_t; F] + \mathbb{E}[M_\infty; \Omega - F] = 0$$

It follows that $\mathbb{E}[M_t; F] = \mathbb{E}[M_\infty; F]$. Since this is true for all $F \in \mathcal{F}_t$, we must have $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$, as required.

$\dashv$

2.7.4 Spaces of Martingales

A martingale M is called an $\mathcal{L}^p$–martingale if and only if $\sup_t |M_t|^p < \infty$, i.e. if and only if M is bounded in $\mathcal{L}^p$. An $\mathcal{L}^2$–martingale is often called *square–integrable.*

The spaces of square–integrable martingales will be important for the definition of the stochastic integral. When we define the stochastic integral, this will be done in several stages. At one stage, we will consider it to be the limit of a sequence of square–integrable martingales. This limit must also be a square–integrable martingale, so the space of martingales must be complete.

Recall that if M is a martingale which is bounded in $\mathcal{L}^2$ is necessarily bounded in $\mathcal{L}^1$, by Hölder's inequality, and is also UI, by Proposition 2.4.6. Thus if M is cádlág, the Martingale Convergence Theorem guarantees the existence of a random variable M_∞ such that $M_t \to M_\infty$ a.s. and in $\mathcal{L}^1$. But does it also converge in $\mathcal{L}^2$? The following theorem proves that this is so:

Theorem 2.7.23 *Suppose that M is a cádlág martingale bounded in $\mathcal{L}^p$, where $p > 1$. Then there is a random variable $M_\infty \in \mathcal{L}^p$ such that*

$$M_t \to M_\infty \quad \text{a.s. and in } \mathcal{L}^p$$

and

$$\sup_{t\geq 0} \mathbb{E}[|M_t|^p] = \lim_{t\to\infty} \mathbb{E}[|M_t|^p] = \mathbb{E}[|M_\infty|^p]$$

Proof: We repeat the argument presented just after the proof of the $\mathcal{L}^p$–inequality (Theorem 2.7.15): Because M is UI (Proposition 2.4.6), it is clear that there is a random variable M_∞ such that $M_t \to M_\infty$ a.s. and in $\mathcal{L}^1$ (by the Martingale Convergence Theorem). Now clearly $|M_t|^p$ is a non–negative submartingale (by Jensen's inequality). Define $M^* = \sup_{t\geq 0} |M_t|$. Using Doob's $\mathcal{L}^p$–inequality, we see that $M^* \in \mathcal{L}^p$, and that $|M_t| \leq M^*$ for each t. Since $M_t \to 0$ a.s., we also have $|M_\infty| \leq M^*$. Thus $|M_t - M_\infty| \leq |M_t| + |M_\infty| \leq 2M^*$, so that $|M_t - M_\infty|^p \leq (2M^*)^p \in \mathcal{L}^1$. By the Dominated Convergence Theorem we thus have $|M_t - M_\infty|^p \to 0$ in $\mathcal{L}^1$, i.e. $M_t \to M_\infty$ in $\mathcal{L}^p$.

Moreover, since $|M_t|^p$ is a submartingale, we see that if $s \leq t$, then

$$\mathbb{E}[|M_s|^p] \leq \mathbb{E}[|M_t|^p]$$

so that $\mathbb{E}[|M_t|^p]$ is increasing (in t). Thus $\sup_{t\geq 0} \mathbb{E}[|M_t|^p] = \lim_{t\to\infty} \mathbb{E}[|M_t|^p]$. But $M_t^p \to M_\infty^p$ a.s., and since $|M_t|^p \leq |M^*|^p \in \mathcal{L}^1$, we also have $\mathbb{E}[|M_t|^p] \to \mathbb{E}[|M_\infty|^p]$ (as $t \to \infty$) by the Dominated Convergence Theorem.

$\dashv$

Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t)$ be a filtered probability space, and suppose that M is a square–integrable martingale. Then $M_t \to M_\infty$ a.s. and in $\mathcal{L}^2$, and $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$. Similarly, given any random variable $X \in \mathcal{L}^2$, we can define a square–integrable martingale M by $M_t = \mathbb{E}[X|\mathcal{F}_t]$. Then $M_t \to M_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$ a.s. Moreover, $\mathbb{E}M_t^2 \leq \mathbb{E}M_\infty^2 \leq \mathbb{E}X^2 < \infty$, so that M_t is a square–integrable martingale with $M_t \to M_\infty$ in $\mathcal{L}^2$ as well.

There is therefore a *correspondence* between square–integrable martingales and square–integrable random $\mathcal{F}_\infty$–measurable variables, which means that the space of square–integrable martingales can be equipped with the same structure as the Hilbert space of square–integrable random variables. This motivates the following definitions:

Definition 2.7.24 (a) $\mathcal{M}^2 = \{\text{martingales } M : M_\infty \in \mathcal{L}^2\}$

(b) $\mathcal{M}_0^2 = \{M \in \mathcal{M}^2 : M_0 = 0 \text{ a.s.}\}$

(c) $c\mathcal{M}^2 = \{M \in \mathcal{M}^2 : M \text{ is continuous a.s.}\}$

(d) $c\mathcal{M}_0^2 = \mathcal{M}_0^2 \cap c\mathcal{M}^2$

We define an inner product on $\mathcal{M}^2$ by

$$\langle M, N\rangle = \mathbb{E}[M_\infty N_\infty]$$

The norm $||\cdot||_{\mathcal{M}^2}$ on $\mathcal{M}^2$ induced by this inner product is

$$||M||_{\mathcal{M}^2} = \left(\mathbb{E}M_\infty^2\right)^{\frac{1}{2}} = ||M_\infty||_2$$

Here the norm on the left is the norm of the martingale, whereas the norm on the right is the usual $\mathcal{L}^2$–norm of the random variable M_∞. We have used the same notation for these norms. Note also that, by Theorem 2.7.23,

$$||M||_2 = \sup_t ||M_t||_2$$

(where again the norm on the left is the $\mathcal{M}^2$–norm and the norm on the right the $\mathcal{L}^2$–norm).

Theorem 2.7.25 *The spaces $\mathcal{M}^2, \mathcal{M}^2_0, c\mathcal{M}^2, c\mathcal{M}^2_0$ are Hilbert spaces.*

Proof: It ought to be clear from the foregoing discussion that $\mathcal{M}^2$ is isomorphic to the Hilbert space $\mathcal{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, and thus itself a Hilbert space. It is trivial to show that all the other spaces inherit the Hilbert space structure from $\mathcal{L}^2$ as well, except possibly for completeness.

We check that $c\mathcal{M}^2_0$ is complete; the proofs for the other spaces are easy adaptations. So let $M^{(n)}$ be a Cauchy sequence of martingales in $c\mathcal{M}^2_0$. Each $M^{(n)}$ converges to some $M^{(n)}_\infty$ a.s. and in $\mathcal{L}^2$. Now because of the way the norm is defined in $\mathcal{M}^2$, it follows that $M^{(n)}_\infty$ is a Cauchy sequence in the complete space $\mathcal{L}^2$, and therefore converges. Thus define

$$\begin{aligned} M_\infty &= \lim_{n\to\infty} M^{(n)}_\infty \qquad \text{(limit in } \mathcal{L}^2) \\ M_t &= \mathbb{E}[M_\infty | \mathcal{F}_t] \end{aligned}$$

It is clear that M so defined is a square–integrable martingale, but we still need to check that M is a.s. continuous and that $M_0 = 0$. Now for each n, $M^{(n)} - M$ is a square–integrable martingale, and so by Theorem 2.7.23 we have

$$\sup_{t\geq 0} \mathbb{E}[(M^{(n)}_t - M_t)^2] = \mathbb{E}[(M^{(n)}_\infty - M_\infty)^2] = ||M^{(n)} - M||^2_{\mathcal{M}^2} \to 0 \quad \text{ as } n \to \infty$$

In particular, $\mathbb{E}M_0^2 = \mathbb{E}(M^{(n)}_0 - M_0)^2 \to 0$, so that $\mathbb{E}M_0^2 = 0$. Thus $M_0 = 0$ a.s.

By Doob's $\mathcal{L}^2$–inequality, we see that $||\sup_t |M^{(n)}_t - M_t|\ ||_2 \leq 2\sup_t ||M^{(n)}_t - M_t||_2 = 2||M^{(n)}_\infty - M_\infty||_2$. Thus $\sup_t |M^{(n)}_t - M_t| \to 0$ in $\mathcal{L}^2$, and thus in probability. It follows that there is a subsequence n_k such that $\sup_t |M^{(n_k)}_t - M_t| \to 0$ a.s. (as $k \to \infty$), and thus that $M^{(n_k)} \to M$ uniformly almost surely on the interval $[0, \infty]$. Since the uniform limit of continuous functions is continuous, M is also continuous a.s.

$\dashv$

2.7.5 Local Martingales

Definition 2.7.26 Let **P** be a property. We say that a stochastic process X_t is locally **P** if and only if there is a sequence of stopping times $\tau_n \uparrow \infty$ s.t. each stopped process $X^{\tau_N}_t$ is **P**. However, *additional* requirements may also have to be met. The sequence τ_n is called a *reducing* or *localizing* sequence for X_t.

Thus we can speak of local martingales, local submartingales, etc. Here, however, we will want to impose some extra requirements:

Definition 2.7.27 A stochastic process X is called a *local martingale* (w.r.t. $(\mathcal{F}_t)_{t\geq 0}$) if and only if there are stopping times $\tau_n \uparrow \infty$ such that each shifted stopped process $X^{\tau_n}_t - X_0$ is a martingale w.r.t. the filtration $(\mathcal{F}_{\tau\wedge t})_{t\geq 0}$. The sequence of stopping times τ_n is called a *localizing sequence*, and is said to *reduce* X.

The reason for specifying that $X^\tau_t - X_0$ be a martingale, rather than just X^τ_t, is that one may encounter circumstances where X_0 may not be integrable. For example, the initial distribution of the process may be weird, but the process is well–behaved from there onwards. In many applications X_0 will be 0.

Note that any martingale is a local martingale: Simply let $\tau_n = n$.

Remarks 2.7.28 There are several reasons why it is often easier to work with local martingales than with martingales:

(1) In the first place, it frequently stops us from having to fuss about integrability, For example, if X is a martingale, and φ is a convex function, then $\varphi(X)$ is always a local submartingale. It will be a submartingale only if each $\mathbb{E}|\varphi(X_t)| < \infty$, which may be difficult to check.

(2) Often we deal with a process X that is only defined on a random time interval $[0, \tau)$. If $\tau < \infty$, then the concept of martingale does not make sense, because $X_t(\omega)$ will be defined only if $t < \tau(\omega)$. It is easy to define a local martingale, however: There are stopping times $\tau_n \uparrow \tau$ such that each X^{τ_n} is a martingale.

(3) The extra generality obtained by working with local martingales is not offset by an increase in complexity of the proofs: Since most of the theorems will be proved by introducing stopping times to reduce the problem to a question about "nice" martingales, the proofs for local martingales are no harder than those for ordinary martingales.

□

Note that not every local martingale X is a martingale, not even if each X_t is integrable.

Exercise 2.7.29 The definition of *local (sub)martingale* differs in some texts. In particular, X_t is often defined to be a local martingale if and only if there is a reducing sequence τ_n such that each $X_{\tau_n \wedge t} - X_0$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. The aim of this exercise is to show that this doesn't affect the definition: $X^\tau_t - X_0$ is an $\mathcal{F}_t$–martingale if and only if it is an $\mathcal{F}_{\tau \wedge t}$–martingale.

(a) Show that if $\mathcal{F}_t$ satisfies the usual conditions, then so does the filtration $\mathcal{F}_{\tau \wedge t}$.

(b) Prove the trivial fact that any $\mathcal{F}_t$–martingale adapted to $\mathcal{F}_{\tau \wedge t}$ is a $\mathcal{F}_{\tau \wedge t}$–martingale.

(c) Suppose that X_t is cádlág and adapted to a filtration $\mathcal{F}_t$ that satisfies the usual conditions. Let τ be a stopping time such that $X^\tau_t - X_0$ is a $\mathcal{F}_{\tau \wedge t}$–martingale. Show that $X^\tau_t - X_0$ is a $\mathcal{F}_t$–martingale.

[Hints: (a) Theorem 2.7.3; (c) Let $F \in \mathcal{F}_s$. Note that $F \cap \{s \leq \tau\} \in \mathcal{F}_{\tau \wedge s}$. Now note that $\mathbb{E}[X^\tau_t - X_0; F] = \mathbb{E}[X^\tau_t - X_0; F \cap \{s \leq \tau\}] + \mathbb{E}[X^\tau_t - X_0; F \cap \{\tau < s\}]$.]

□

It must be stressed that, though every martingale is a local martingale, the converse is not true. We shall see some examples once we've covered stochastic integration. Nevertheless, we sometimes will want to know when a local martingale is actually a martingale. The following will aid in this regard:

Proposition 2.7.30 *Let $p \in [1, \infty)$ and let X be a local $\mathcal{L}^p$–martingale, reduced by τ_n. If $\{|X^{\tau_n}_t|^p : n \in \mathbb{N}\}$ is UI for each $t \geq 0$, then X is an $\mathcal{L}^p$–martingale.*

Proof: Note that $|X_0| = |X_{0\wedge\tau_n}|$ for each n. Thus $X_0 \in \mathcal{L}^p$ (by the UI assumption with $t = 0$). It follows that each $X_t^{\tau_n}$ is an $\mathcal{L}^p$–martingale, and thus we have $X_t^{\tau_n} \to X_t$ a.s. as $n \to \infty$. Moreover, the UI assumption implies that convergence is in $\mathcal{L}^p$ as well (Theorem 2.4.5). It follows that $\mathbb{E}[X_t^{\tau_n}|\mathcal{F}_s] \to \mathbb{E}[X_t|\mathcal{F}_s]$ in $\mathcal{L}^p$. Putting the pieces together yields the result.

$\dashv$

Corollary 2.7.31 *If X_t is a local martingale and if $\mathbb{E}[\sup_{0\le s\le t} |X_s|] < \infty$ for each $t \ge 0$, then X_t is a martingale.*

Exercise 2.7.32 Prove the preceding corollary.

$\square$

Proposition 2.7.33 *If X_t is a local martingale, and if τ is a stopping time, then X_t^τ is a local martingale, and any sequence that reduces X_t will also reduce X_t^τ.*

Proof: Suppose that $\sigma_n \uparrow \infty$ reduces X_t. Then $X_t^{\sigma_n} - X_0$ is a martingale, and so $X_t^{\sigma_n\wedge\tau} - X_0 = (X_t^{\sigma_n} - X_0)^{\tau_n}$ is a martingale (Stopped martingales are martingales). The result follows easily.

$\dashv$

The following proposition provides a nice class of stopping times which form reducing sequences for *continuous* local martingales:

Proposition 2.7.34 *Suppose that X_t is a continuous local martingale. Let $\tau_n = \inf\{t : |X_t| > n\}$. Then τ_n reduces X. Indeed, let $\tau_n' \le \tau_n$ be a sequence of stopping times with $\tau_n' \uparrow \infty$. Then τ_n' reduces X.*

Proof: That τ_n are stopping times we leave to the next exercise. Since stopped local martingales are local martingales, each $X_t^{\tau_n'}$ is a local martingale. Moreover, since X_t is continuous, we have

$$\mathbb{E} \sup_{0\le s\le t} |X_t^{\tau_n'}| \le n < \infty \qquad \text{for all } t$$

which implies that $X_t^{\tau_n'}$ is a martingale (by Corollary 2.7.31), and thus that τ_n' reduces X.

$\dashv$

Exercise 2.7.35 Prove that if X_t is continuous, then $\tau_n = \inf\{t : |X_t| > n\}$ is a stopping time.
[Hint: By continuity, $\{\tau_n > t\} = \bigcap_{q\le t, q\in\mathbb{Q}} \{|X_q| \le n\}$]

$\square$

Chapter 3

Prelude to Stochastic Integration

3.1 The Riemann–Stieltjes Integral: Motivation and Definition

Example 3.1.1 It is well–known that area under a continuous curve $g(t)$ over the interval $[a,b]$ is given by the Riemann integral $\int_a^b g(t)\,dt$. In elementary calculus courses, this integral is defined as a limit of sums. Roughly, one *partitions* the interval $[a,b]$ into equally spaced points $a = t_0 < t_1 < t_2 < \cdots < t_n = b$, and chooses an element $t_k^* \in [t_{k-1}, t_k]$ for each $k = 1, \dots, n$. One then considers the *Riemann sums*

$$\sum_{k=1}^{n} g(t_k^*)\ (t_k - t_{k-1}) = \sum_{k=1}^{n} g(t_k^*)\Delta t$$

The Riemann integral $\int_a^b g(t)\ dt$ is defined to be limit of these Riemann sums as $\Delta t \to 0$, assuming this limit exists.

This description of the Riemann integral is not completely precise — you will learn how to make it precise shortly.

$\square$

Example 3.1.2 Suppose that X is a random variable with values in the interval $[a,b]$, and with distribution function $F : \mathbb{R} \to [0,1] : x \mapsto \mathbb{P}(X \leq x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be an arbitrary continuous function. To estimate $\mathbb{E}[g(X)]$, one can proceed as follows: Partition the interval $[a,b]$ into subintervals with endpoints $a = x_0 < x_1 < \cdots < x_n = b$ and choose $x_k \in [x_{k-1}, x_k]$ for $k = 1, \dots, n$. If $\Delta_k x := x_k - x_{k-1}$ is sufficiently small, then the function g, being continuous, is almost constant on the interval $[x_{k-1}, x_k]$, and has value $\approx g(x_k^*)$ on that interval. We can now approximate X by a discrete random variable $\tilde{X}$ with values in $\{x_1^*, \dots, x_n^*\}$ defined by

$$\tilde{X} := x_k^* \qquad \Longleftrightarrow \qquad X \in (x_{k-1}, x_k]$$

So we have

$$\mathbb{E}[g(X)] \approx \mathbb{E}[g(\tilde{X})] = \sum_{k=1}^{n} g(x_k^*)\mathbb{P}(\tilde{X} = x_k^*)$$

But $\tilde{X} = x_k^*$ just when $X \in (x_{k-1}, x_k]$, i.e.

$$\mathbb{P}(\tilde{X} = x_k^*) = \mathbb{P}(x_{k-1} < X \leq x_k) = \mathbb{P}(X \leq x_k) - \mathbb{P}(X \leq x_{k-1}) = F(x_k) - F(x_{k-1})$$

Thus

$$\mathbb{E}[g(X)] \approx \sum_{k=1}^{n} g(x_k^*) \left(F(x_k) - F(x_{k-1})\right) = \sum_k g(x_k^*)\ \Delta_k F$$

The limit of these sums, as $\Delta x \to 0$ is a *Riemann–Stieltjes* integral, and denoted $\int_a^b g(x)\ dF(x)$

□

Example 3.1.3 Suppose that at time t, you own $\theta(t)$–many shares, and that the share price at time t is $S(t)$. What is your gain/loss over a time period $[0,T]$?

At time t, you have $\theta(t)$ shares. Between times t and $t+\Delta(t)$, the share price changes by $\Delta S(t) := S(t+\Delta t) - S(t)$, and so your gainover that period is $\approx \theta(t)\Delta S(t)$. To approximate your total gain over the period $[0,T]$, partition the interval $0 = t_0 < t_1 < \cdots < t_n = T$, to obtain

$$\text{Gain} \approx \sum_{k=1}^{n} \theta(t_k)\Delta S(t_k)$$

Intuitively, the approximation becomes more and more accurate as $\Delta t \to 0$, so the gain of the portfolio θ is, roughly, the limit $\int_0^T \theta(t)\ dS(t) = \lim_{\Delta t \to 0} \sum_{k=1}^n \theta(t_k)\Delta S(t_k)$ provided this limit exists[1].

□

We have now seen several situations where we need to determine the limit of sums $\sum_{k=1}^n f(t_k)\ dG(t_k)$, where the limit is taken as the $\Delta t \to 0$. The description so far has been intuitive and imprecise. It is now time to introduce some rigour.

Let f, G be real–valued functions defined and bounded on an interval $[a,b]$. A *partition* P of $[a,b]$ is a a finite ordered set $\{a = t_0 < t_1 < t_2 < \cdots < t_n = b\}$. The *size* of such a partition is denote $\sigma(P)$, and defined by

$$\sigma(P) := \max_k (t_k - t_{k-1})$$

A *tagged partition* is a partition P together with a choice $t_k^* \in [t_{k-1}, t_k]$ for each $k = 1, \ldots, n$. Tagged partitions will be indicated by a $*$, i.e. if P is a partition, then P^* dentes an associated tagged partition.

With each tagged partition, we can associate a Riemann–Stieltjes sum (abbreviated RS sum)

$$S(P^*, f, G) := \sum_{k=1}^{n} f(t_k^*)\ (G(t_k) - G(t_{k-1}) = \sum_{k=1}^{n} f(t_k^*)\ \Delta_k G$$

The Riemann–Stieltjes integral (abbreviated RS integral) $\int_a^b f\ dG$ should be the limit of the RS sums, over all tagged partitions P^*, as $\sigma(P) \to 0$. To be precise, we say

$$\lim_{\sigma(P)\to 0} S(P^*, f, G) = L \text{ exists}$$

if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$|S(P, f, G) - L| < \varepsilon \qquad \text{whenever} \qquad \sigma(P) < \delta$$

[1]Unfortunately, for most reasonable models of asset dynamics $S(t)$, this limit does not exist; a *stochastic* integral must be used...

Then we define

$$\int_a^b f\, dG := \lim_{\sigma(P)\to 0} S(P^*, f, G)$$

provided this limit exists, and say that f is Riemann–Stieltjes integrable with respect to G on $[a, b]$.

When the function G is the identity function $G(t) = t$, the Riemann–Stieltjes integral is just the ordinary Riemann integral.

With each partition $\{a = t_0 < t_1 < \cdots < t_n = b\}$ it is possible to associate three natural tagged partitions, namely those having tags equal to the left endpoint, right endpoint and midpoint of each interval. This yields:

- The lefthand RS sum $\sum_k f(t_{k-1})\Delta_k G$;
- The righthand RS sum $\sum_k f(t_k)\Delta_k G$;
- The symmetric RS sum $\sum_k f(\frac{t_{k-1}+t_k}{2})\Delta_k G$.

If f is RS integrable w.r.t. G, then each of these sums must converge as $\sigma(P) \to 0$, and all to the same limit.

Remarks 3.1.4 A slightly different definition uses *Darboux sums* rather than RS sums. Given real–valued functions f, G defined and bounded on an interval $[a, b]$, and a partition $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$, let the upper and lower Darboux sums be defined by

$$U(P, f, G) := \sum_{k=1}^{n} \sup\{f(t) : t \in [t_{k-1}, t_k]\} \cdot (G(t_k) - G(t_{k-1}))$$
$$L(P, f, G) := \sum_{k=1}^{n} \inf\{f(t) : t \in [t_{k-1}, t_k]\} \cdot (G(t_k) - G(t_{k-1}))$$

If f is continuous on $[a, b]$, it attains its supremum and infimum on each subinterval, i.e. we can choose $t_k^{\max}, t_k^{\min} \in [t_{k-1}, t_k]$ such that

$$f(t_k^{\max}) = \sup\{f(t) : t \in [t_{k-1}, t_k]\} \qquad f(t_k^{\min}) = \inf\{f(t) : t \in [t_{k-1}, t_k]\}$$

If $P^{*\max}, P^{*\min}$ are the tagged partitions given by $a = t_0 < \cdots < t_n = b$ and the tags $t_k^{\max}, t_k^{\min}$ respectively, then it is easy to see that

$$U(P, f, G) = S(P^{*\max}, f, G) \qquad L(P, f, G) = S(P^{*\min}, f, G)$$

i.e. the Darboux sums give the most extreme values of the Riemann–Stieltjes sums for any given partition. However, the Darboux sums may differ from Riemann–Stieltjes sums if f is not continuous.

The Riemann–Stieltjes sums may be defined even when f, G are Banach space valued, however, whereas the Darboux sums, being dependent on sup's and inf's, make sense for real–valued functions only.

□

Now that we have defined $\int_a^b f\ dG$ as a certain limit, the first question that begs our attention is the following: Under what conditions *does* this limit exist?

Here are a few instances where the answer is obvious:

Example 3.1.5 (1) If $G = \mathrm{id}$ is the identity function on $[a, b]$ (i.e. $G(t) = t$ for all t), then $\int_a^b f\ dG$ is just the ordinary Riemann integral of f with respect to t, and will exist whenever f is, for example, piecewise continuous on $[a, b]$.

(2) If $f(t) = 1$ for all $t \in [a, b]$, then

$$S(P^*, f, G) = \sum_k (G(t_k) - G(t_{k-1})) = G(b) - G(a)$$

for all tagged partitions P, and hence $\int_1^b 1\ dG = G(b) - G(a)$ for any function G.

(3) If G is differentiable, with $G'(t) = g(t)$, and $\{a = t_0 < t_1 < \cdots < t_n = b$, there is by the Mean Value theorem a $t_k^* \in [t_{k-1}, t_k]$ such that

$$G(t_k) - G(t_{k-1}) = g(t_k^*)(t_k - t_{k-1})$$

Thus

$$S(P^*, f, G) = \sum_k f(t_k^*)g(t_k^*)(t_k - t_{k-1}) = S(P, fg, \mathrm{id})$$

and so

$$\int_a^b f\ dG = \int_a^b f(t)g(t)\ dt = \int_a^b f(t)G'(t)\ dt$$

is again just an ordinary Riemann integral. It will exist if, for example, f, G' are both piecewise continuous on $[a, b]$.

□

Example 3.1.6 Here is a simple example which shows when the Riemann–Stieltjes integral may fail to exist. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(t) = \begin{cases} 0 \text{ if } t < \dfrac{1}{2} \\ 1 \text{ if } t \geq \dfrac{1}{2} \end{cases}$$

Coinsider partitions $\{0 = t_0 < t_1 < \cdots < t_n = 1\}$ of $[0, 1]$ with the property that the point $\frac{1}{2}$ is *not* one of the endpoints t_k. For each such partition, there is a unique k_0 such that $t_{k_0-1} < \frac{1}{2} \leq t_{k_0}$. Then

$$\Delta_k f = \begin{cases} 0 \text{ if } k \neq k_0 \\ 1 \text{ if } k = k_0 \end{cases}$$

Looking at the left– and righthand RS sums, we see that

$$\sum_k f(t_{k-1})\Delta_k f = 0 \qquad \sum_k f(t_k)\Delta_k f = 1$$

for all partitions, no matter how fine. Hence $\int_0^1 f\ df$ does not exist.

□

This argument can easily be adapted to prove that $\int_a^b f\ \ dG$ cannot exist if f, G share a common point of discontinuity in $[a, b]$.

3.2 Functions of Finite Variation

Suppose that f, G are bounded on $[a,b]$. Let M be a bound for f, so that $|f(t)| \leq M$ for all $t \in [a,b]$. Then

$$\sum_{k=1}^{n} f(t_k^*)(G(t_k) - G(t_{k-1})) \leq M \sum_{k=1}^{n} |G(t_k) - G(t_{k-1})|$$

For $\int_a^b f\,dG$ to exist for a variety of functions f, it is therefore necessary that the quantity

$$\lim_{\sigma(P)\to 0} \sum_{k=1}^{n} |G(t_k) - G(t_{k-1})|$$

does not get out of hand, i.e. does not diverge to $+\infty$ as partition meshes get smaller and smaller. This motivates the following definition:

Definition 3.2.1 Let $G : [a,b] \to \mathbb{R}$, and let $P = \{a = t_0 < t_1 < \cdots < t_n\}$ be a partition of $[a,b]$. Define the *variation* of G on P by

$$V_P(G;[a,b]) = \sum_{k=1}^{n} |G(t_k) - G(t_{k-1})|$$

Define the (total) *variation* of G on $[a,b]$ by

$$V(G;[a,b]) = \sup_P V_P(G;[a,b])$$

where the supremum is taken over all partitions of $[a,b]$.
The function G is said to be of *finite variation* on $[a,b]$ provided that $V(G;[a,b]) < \infty$.

$\square$

A function is said to be *of locally variation* if its variation on every compact interval is finite.

If P, Q are partitions of $[a,b]$, we say that Q is *finer* than P iff $P \subseteq Q$. This simply means that every subdivision of P is a union of subdivisions of Q. Using the triangle inequality, it is easy to see that if Q is finer than P, then $V_P(G) \leq V_Q(G)$, and thus that $V_P(G)$ increases as P gets finer and finer. Thus $\lim_{\sigma(P)\to 0} V_P(G) = \sup_P V_P(G) = V(G)$.

The quantity $V(G;[a,b])$ is very important: If $V(G;[a,b]) = \infty$, we may well expect that $\int_a^b f\,dG$ does not exist for a variety of functions f.

The following properties are easy to prove:

Proposition 3.2.2 *(a) If f is increasing on $[a,b]$, then $V(f;[a,b]) = f(b) - f(a)$. Thus any monotone function is of locally finite variation.*

(b) If $s \leq t$, then $V(f;[a,s]) \leq V(f;[a,t])$.

(c) If f is differentiable, then $V(f;[a,b]) = \int_a^b |f'(s)|\,ds$. Thus any continuously differentiable function is of locally finite variation.

(d) $V(f;[a,b]) \geq |f(b) - f(a)|$.

(e) If $\alpha \in \mathbb{R}$ and if f, g are of (locally) finite variation, then so are $f + g$ and αf, and

$$V(f; [a, t]) \leq V(f; [a, t]) + V(g; [a, t]) \qquad V(\alpha f; [a, t]) = |\alpha| V(f; [a, t])$$

Hence the family of functions of bounded variation on a compact interval is a vector space.

(f) If $a < b < c$, then $V(f; [a, c]) = V(f; [a, b]) + V(f; [b, c])$.

□

Exercise 3.2.3 (a) Prove the preceding proposition.

(b) Calculate the variation of $\sin(x)$ over $[0, 2\pi]$ directly from the definition. Verify your answer by using (c) of the preceding proposition.

(c) Find all functions f on $[a, b]$ with the property that $V(f; [a, b]) = 0$.

□

Can you think of any bounded functions which are not of finite variation on a compact interval? At first glance, probably not. If you understood the previous exercise, you will know that the variation is essentially the sum of the sizes of all the "bounces", i.e. the sum of the distances from each local minimum to the next local maximum, and from each local maximum to the next local minimum.

Example 3.2.4 Let $f(x) = x^2$. On [-2,3], f first bounces down from 4 to 0, and then bounces up from 0 to 9. The variation $V_f(-2, 3)$ is therefore equal to $4 + 9 = 13$.

□

This shows that a function is of bounded variation if it is *not too bouncy*, i.e. if the sum of all the bounces does not add up to $+\infty$.

Exercise 3.2.5 Consider the functions

$$f(x) = \begin{cases} x \sin(\frac{1}{x}) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases} \qquad g(x) = \begin{cases} x^2 \sin(\frac{1}{x}) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

(a) Show that f, g are continuous.

(b) Show that f is not of bounded variation on $[-\pi, \pi]$.

(c) Sketch a graph of f to see why it isn't of finite variation. Also sketch a graph of g.

(d) Show that g is of finite variation on $[-\pi, \pi]$.

[Hints: (b) $f(x)$ reaches a local maximum for values $x = \frac{2}{(4n+1)\pi}$ and a local minimum for values $x = \frac{2}{(4n+3)\pi}$. Thus

$$\begin{aligned} V_f(-\pi,\pi) &= \frac{2}{\pi}\sum_{n=0}^{\infty}\left[\frac{1}{4n+1}+\frac{1}{4n+3}+\frac{1}{4n+3}+\frac{1}{4n+5}\right] \\ &\geq \frac{2}{4\pi}\sum_{n=1}^{\infty}\frac{1}{n} \\ &= +\infty \end{aligned}$$

because the harmonic series diverges.
(d) $\sum_n \frac{1}{n^2}$ converges.]

$\square$

Suppose that f is of finite variation on a compact interval $[a,b]$. For $s,t \in [a,b]$ define $V_f(s,t) := V(f;[s,t])$.

Proposition 3.2.6 *If f is continuous and of finite variation on a compact interval $[a,b]$ then $V_f(a,t)$ is continuous in t.*

Proof: Since $V_f(a,t)$ is increasing in t, we just need to rule out jumps. We first show that $V_f(a,t)$ is left–continuous at $t \in [a,b]$, i.e. that $V_f(a,s) \uparrow V_f(a,t)$ as $s \uparrow t$. Since

$$V_f(a,t) = V_f(a,s) + V_f(s,t)$$

we need only show that $V_f(s,t) \to 0$ as $s \uparrow t$. Now if this is not the case, then there is $\delta > 0$ such that $V_f(s,t) > \delta$ for all $s < t$. Let $s_1 < t$ be arbitrary. Choose a partition $P_1 = \{s_1 = t_0^1 < t_1^1 < \cdots < t_n^1\}$ of $[s_1,t]$ such that

$$\sum_{k=1}^{n} |f(t_k^1) - f(t_{k-1}^1)| > \delta$$

Since f is continuous at t we may choose s_2 sufficiently close to t such that the inequality still holds if we replace the last term $|f(t) - f(t_{n-1}^1)|$ in the above sum by $|f(s_2) - f(t_{n-1}^1)|$. This proves that there is s_2 such that $V_f(s_1,s_2) > \delta$, which shows that

$$V_f(a,t) \geq V_f(s_1,t) = V_f(s_1,s_2) + V_f(s_2,t) > 2\delta$$

Now repeat the entire argument, replacing s_1 by s_2, to find an s_3 very close to t so that $V_f(s_2,s_3) > \delta$. We then have

$$V_f(a,t) \geq V_f(s_1,t) = V_f(s_1,s_2) + V_f(s_2,s_3) + V_f(s_3,t) > 3\delta$$

Next find s_4 sufficiently close to t such that $V_f(s_3,s_4) > \delta$, to prove that $V_f(a,t) > 4\delta$, etc. It follows that $V_f(a,t) = \infty$, contradiction. Hence $V_f(a,t)$ is left–continuous (in t).

A similar argument will show that $V_f(a,t)$ is right–continuous.

$\dashv$

The following result is tremendously useful:

Proposition 3.2.7 *A function f is of finite variation over a compact interval if and only if f can be represented as the difference $f = g - h$ of two increasing functions g, h. Moreover, if f is continuous, then we can choose g, h to be continuous as well.*

Proof: It is easy to see that if g, h are increasing, then $g - h$ is of locally finite variation. (Exercise!) Now assume that f is of finite variation on $[a, b]$. Define

$$g(t) = \frac{1}{2}\left[V_f(a,t) + f(t)\right] \qquad h(t) = \frac{1}{2}\left[V_f(a,t) - f(t)\right]$$

It is clear that $g - h = f$, and that if f is continuous, then so are g, h (because $V_f(a,t)$ is continuous in t). So it remains to show that g, h are increasing. But if $s \leq t$, then

$$\begin{aligned} g(t) - g(s) &= \frac{1}{2}\left[V_f(a,t) - V_f(a,s) + f(t) - f(s)\right] \\ &= \frac{1}{2}\left[V_f(s,t) + f(t) - f(s)\right] \\ &\geq 0 \end{aligned}$$

because $V_f(s,t) \geq |f(t) - f(s)|$. Thus $g(t) \geq g(s)$, and so g is increasing. A similar argument holds for h.

⊣

Exercise 3.2.8 Show that if $f(x)$ is of finite variation, it is discontinuous at at most countably many x.
[Hint: Since f is the difference of two increasing functions, it suffices to show that an increasing function can have at most countably many points of discontinuity. Define

$$A_n = \{x : f \text{ is discontinuous at } x \text{ with jump } > \frac{1}{n}\}$$

If f has uncountably many jumps, then one of the A_n is uncountable.]

□

We now prove a criterion which guarantees the existence of $\int_a^b f\, dG$:

Theorem 3.2.9 *If f is continuous and G is of bounded variation over $[a, b]$, then $\int_a^b f\, dG$ exists.*

Proof: Since f is continuous on the compact interval $[a, b]$, it is uniformly continuous. Thus, given $\varepsilon > 0$, we may choose $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

Suppose that

$$P = \{a = t_0 < t_1 < \cdots < t_n = b\}$$

is a partition of $[a, b]$ with $\sigma(P) < \delta$, and choose $u_k, l_k \in [t_{k-1}, t_k]$ such that

$$f(u_k) = \sup\{f(t) : t \in [t_{k-1}, t_k]\} \qquad f(l_k) = \inf\{f(t) : t \in [t_{k-1}, t_k]\}$$

u_k, l_k exist because a continuous function has a maximum and a minimum on any compact set. Let $P^{*\max}, P^{*\min}$ be the associated tagged partitions, with tags u_k, l_k respectively. If P^* is any other tagged partition based on P, we clearly have

$$S(P^{*\min}, f, G) \leq S(P^*, f, G) \leq S(P^{*\max}, f, G)$$

so to prove that f is RS–integrable w.r.t. G it suffices to show that if $\Delta(P, f, G) := S(P^{*\max}, f, G) - S(P^{*\min}, f, G)$, then $\lim_{\sigma(P)\to 0} \Delta(P, f, G) = 0$.

Note that

$$S(P^{*\max}, f, G) = \sum_k f(u_k)\ \Delta_k G \qquad S(P^{*\min}, f, G) = \sum_k f(l_k)\ \Delta_k G$$

and thus

$$\Delta(P, f, G) = \sum_k (f(u_k) - f(l_k))\ \Delta_k G$$

Hence

$$|\Delta(P, f, G)| \leq \varepsilon \sum_k |\Delta_k G| \leq \varepsilon V(G; [a, b])$$

Since $V(G; [a, b])$ is finite, by assumption, it follows that $\Delta(P, f, G)$ can be made arbitrarily small by choosing P to be sufficiently fine.

⊣

More can be said: If f is merely piecewise continuous on $[a, b]$, and if G is of finite variation such that f, G have no common point of discontinuity, then $\int_a^b f\ dG$ exists.

The converse of Theorem 3.2.9 is not true, however: $\int_a^b f\ dG$ may exist even if G is *not* of finite variation. The following theorem shows why: If f is continuous, and of bounded variation, and if G is continuous, but possibly of infinite variation, then $\int_a^b f\ dG$ exists, because $\int_a^b G\ df$ exists (by Theorem 3.2.9).

Proposition 3.2.10 (Integration by Parts)
If f is integrable with respect to g, then g is integrable with respect to f, and

$$\int_a^b g\ df = f(b)g(b) - f(a)g(a) - \int_a^b f\ dg$$

Note that if f, g are differentiable, then this is just the ordinary integration–by–parts formula of first–year calculus.

Proof: Let $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a partition of $[a, b]$, and let $t_k^* \in [t_{k-1}, t_k]$ for $k = 1 \ldots n$. Note that $Q = \{a = t_0^* < t_1^* < \cdots < t_n^* \leq t_{n+1}^* = b\}$ is also a partition of $[a, b]$, and that as P gets finer, so does Q (and vice versa). Also note that $t_k \in [t_k^*, t_{k+1}^*]$, for $k = 1, \ldots, n$. Now

$$\begin{aligned}\sum_{k=1}^n f(t_k^*)[(g(t_k) - g(t_{k-1})] &= \sum_{k=1}^n g(t_k)[f(t_k^*) - f(t_{k+1}^*)] - f(a)g(a) + f(b)g(b) \\ &= f(b)g(b) - f(a)g(a) - \sum_{k=1}^n g(t_k)[f(t_{k+1}^*) - f(t_k^*)]\end{aligned}$$

so the result follows by taking limits as $\sigma(P) \to 0$.

⊣

Nevertheless, we do have the following partial converse to Theorem 3.2.9. It depends on the Banach–Steinhaus Theorem[2], a result in Banach space theory. Omit the proof if, as I expect, you don't know this theorem.

Theorem 3.2.11 *If $\int_a^b f\, dG$ exists for all continuous f on $[a, b]$, then G is of bounded variation on $[a, b]$.*

Proof: Let $\mathcal{C}[a, b]$ be the Banach space of continuous functions $f : [a, b] \to \mathbb{R}$, equipped with the supremum norm. Assume that $G : [a, b] \to \mathbb{R}$ has the property that $\int_a^b f\, dG$ exists for all $f \in \mathcal{C}[a, b]$. We shall prove that G is of bounded variation. Let P_n be a sequence of partitions of $[a, b]$ with $||P_n|| \to 0$.

If $P_n = \{a = t_0 < t_1 < \cdots < t_m = b\}$, define a linear operator $T_n : \mathcal{C}[a, b] \to \mathbb{R}$ by

$$T_n(f) = \sum_{k=1}^{m} f(t_{k-1})\ \Delta_k G$$

(where $\mathbb{R}$ is a Banach space equipped with the usual (absolute value) norm.) Let $f_n \in \mathcal{C}[a, b]$ be such that $f_n(t_{k-1}) = \mathrm{sign}\Delta_k G$, with $||f_n|| = 1$. Then

$$T_n(f_n) = \sum_{k=1}^{m} |\Delta_k G|$$

and so

$$||T_n|| \geq \sum_{k=1}^{m} |\Delta_k G|$$

It follows that

$$\sup_n ||T_n|| \geq V_G(a, b)$$

Now by assumption, $\lim_n T_n(f) = \int_a^b f\ dG$ exists for each $f \in \mathcal{C}[a, b]$, and hence the set $\{T_n(h) : n \in \mathbb{N}\}$ is bounded. By the Banach–Steinhaus Theorem, the set $\{||T_n|| : n \in \mathbb{N}\}$ is also bounded, i.e. $\sup_n ||T_n|| < \infty$. It follows that $V_G(a, b) < \infty$, i.e. that G is of bounded variation.

⊣

The following result may be found in any good (advanced) text on Real Analysis.

Theorem 3.2.12 *A function locally of bounded variation is differentiable almost everywhere.*

□

The converse is false: The function $f(x) = x \sin(\frac{1}{x})$ is everywhere continuous, and differentiable almost everywhere (except at $x = 0$), yet not of bounded variation.

[2] Also called the Principle of Uniform Boundedness.

3.3 The Lebesgue–Stieltjes Integral

It is not hard to generalize the Riemann–Stieltjes integral to a Lebesgue–Stieljes integral. Recall that Lebesgue measure λ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which assigns to each interval (a,b) its length, e.g. $\lambda(a,b] = b-a$. Similarly, if G is a right–continuous increasing function, we can define a measure μ_G on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which assigns to each interval (a,b) the measure $\mu_F(a,b] = G(b) - G(a)$. This is called the Lebesgue–Stieltjes measure of G. Of course, we've left out a lot of details — We have to check that μ_G is countably additive on the algebra of left half–open intervals, so that Carathéodory's Extension Theorem applies, etc. But you get the idea.

If F is of bounded variation, then F can be represented as a difference of two increasing functions $F = G - H$. We can then define the Lebesgue–Stieltjes integral of a function $u(t)$ with respect to $F(t)$ over $B \in \mathcal{B}(\mathbb{R})$ by

$$\int_B u \, dF = \int_B u \, d\mu_G - \int_B u \, d\mu_H$$

3.4 Stieltjes Integration of Stochastic Processes

We shall attempt to define the stochastic integral

$$\int_0^T H_t(\omega) \, dX_t(\omega)$$

for adapted processes H_t, X_t. Of course, we assume that we work in a probability space equipped with a filtration $(\mathcal{F}_t)_t$, and that H_t, X_t are adapted to this filtration. If X_t is (almost surely) of bounded variation on $[0,T]$, and if H_t is a.s. continuous, then it follows from the previous section that we can define the stochastic integral pathwise as a random variable whose values are Riemann–Stieltjes integral, one for each $\omega \in \Omega$. Even if H_t is not a.s. continuous, but bounded and adapted, we will be able to define the stochastic integral a random variable whose values are Lebesgue–Stieltjes integrals. This means that we have now successfully defined $\int_0^T H_t \, dX_t$ in the case where H_t is a.s. bounded and adapted, and X_t is a Poisson process (see **Remarks** below), for example, because Poisson processes are increasing, and thus of bounded variation.

Since $H_t(\omega)$ is Riemann–Stieltjes integrable with respect to X_t if and only if X_t is Riemann–Stieltjes integrable with respect to H_t ("Integration by Parts") we have a valid definition of $\int_0^T H_t \; dX_t$ whenever H_t, X_t are continuous, and one of them is of bounded variation.

For example, by using the integration–by–parts formula, we can interpret $\int_0^\pi \sin(t) \, dW_t(\omega)$ to be $-\int_0^\pi W_t(\omega) \, d\sin(t) = -\int_0^\pi W_t(\omega) \cos t \, dt$, an ordinary Riemann integral for each $\omega \in \Omega$.

Remarks 3.4.1 Recall that a random variable Y is Poisson with parameter α if it takes only non–negative integers as values, and if

$$\mathbb{P}(Y = k) = \frac{\alpha^k}{k!} e^{-\alpha}$$

The mean and variance of Y are both easily shown to be equal to α. For example, Y can be interpreted as the number of times a certain event has occurred per unit time, where the average rate of occurrence is α.

$(X_t : t \geq 0)$ is Poisson process provided that

(1) $X_0 = 0$

(2) For $0 \leq s < t < \infty$, $X_t - X_s$ is a Poisson random variable with mean $\alpha(t-s)$

(3) For $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, the set of random variables $\{X_{t_{k+1}} - X_{t_k} : k = 1, \ldots, n\}$ is independent.

One can think of X_t as the number of times a certain event has occurred by time t, if they occur independently at an average rate of α per unit time. Any Poisson process has a version with right–continuous paths, and these paths are almost surely constants, except for upward jumps of size 1. Only finitely (a.s.) many jumps will occur in any bounded time interval. Closely associated with a Poisson process is a family T_k of stopping times, where T_k is time between the k^{th} and $(k+1)^{\text{th}}$ jump. These T_k are exponentially distributed, with $\mathbb{P}(T_k \leq t) = 1 - e^{-\alpha t}$.

It is easy to see that $M_t = X_t - \alpha t$ is a martingale of bounded variation, but M_t is not continuous.

□

We can, for example, model a stock price process with jumps using the sum of a Brownian motion and a Poisson process, a so–called *jump diffusion*.

Let's see what happens if we try to interpret the more interesting integral

$$I = \int_0^T W_t \, dW_t$$

as a Riemann–Stieltjes integral. If the standard rules of calculus apply, we would get

$$I = \frac{1}{2}(W_T^2 - W_0^2) = \frac{1}{2}W_T^2$$

Taking expectations, we'd therefore get

$$\mathbb{E}[I] = \frac{1}{2}T$$

because $\mathrm{Var}(W_T) = T$. That's what we'd get *if the usual rules of calculus apply*[3].

Taking a Riemann–Stieltjes approach, let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0, T]$. We now make two choices for the $t_k^* \in [t_{k-1}, t_k]$. First, let $t_k^* = t_{k-1}$ be the leftmost point of the interval $[t_{k-1}, t_k]$. We then approximate the stochastic integral by

$$I \approx \sum_k W(t_{k-1}, \omega) \Delta_k W(\omega)$$

where $\Delta_k W(\omega) = W(t_k, \omega) - W(t_{k-1}, \omega)$. Taking expectations, we obtain

$$\mathbb{E}[I] \approx \sum_k \mathbb{E}[W_{t_{k-1}} \Delta_k W] = \sum_k \mathbb{E}[W_{t_{k-1}}] \cdot \mathbb{E}[\Delta_k W] = 0$$

because, by definition of Brownian motion, $W_{t_{k-1}}$ is independent of $\Delta_k W$.

[3] But they don't...

Now we make a second choice of t_k^*: We choose it to be the rightmost point, i.e. $t_k^* = t_k$. Note that

$$W(t_k,\omega)\Delta_k(\omega) = \Delta_k(\omega)^2 + W(t_{k-1},\omega)\Delta_k W$$

If we approximate the stochastic integral by

$$I \approx \sum_k W(t_k,\omega)\Delta_k W(\omega)$$

and take expectations, we obtain

$$\mathbb{E}[I] \approx \sum_{\mathbb{E}}[(\Delta_k)^2] = \sum_k (t_k - t_{k-1}) = T$$

because $\Delta_k W$ is $N(0, t_k - t_{k-1})$–distributed (by definition of Brownian motion).

It immediately follows that W_t is (a.s.) not Riemann–Stieltjes integrable with respect to itself, for if it were, the value of I would be independent of our choice of the family t_k^*. This kills our naive attempt to interpret stochastic integrals pathwise as Riemann–Stieltjes integrals.

Exercise 3.4.2 Show that if we choose t_k^* to be the midpoint of $[t_{k-1}, t_k]$, then we'd get $\mathbb{E}[I] \approx \frac{1}{2}T$, which is what ordinary calculus predicts.

□

The next exercise proves an important result:

Exercise 3.4.3 Show that if M_t is an $\mathcal{F}_t$–martingale, then

$$\mathbb{E}\left[(M_t - M_s)^2|\mathcal{F}_s\right] = \mathbb{E}\left[M_t^2 - M_s^2|\mathcal{F}_s\right]$$

This equation is referred to as the *orthogonality of martingale increments* because it depends on the fact that $\mathbb{E}\left[M_s(M_t - M_s)|\mathcal{F}_s\right] = 0$. This means that M_s and $M_t - M_s$ are orthogonal in the Hilbert space $\mathcal{L}^2$.

□

The following result hits the final nail in the coffin:

Proposition 3.4.4 *If X_t is a continuous (a.s.) martingale locally of bounded variation on $[0,T]$, then X_t is constant a.s. on $[0,T]$.*

Proof: Without loss of generality, we may assume that $X_0 = 0$, and that the underlying filtration is complete. (We can always consider the martingale $X_t - X_0$ if needs be, and augment the filtration without changing the martingale property of X.) Let $V_X(t,\omega)$ be the variation of $X_t(\omega)$ on $[0,t]$. Given an arbitrary $K > 0$, define a stopping time τ by

$$\tau = \inf\{t > 0 : V_X(t) > K\}$$

We now show that $Y_t = X_t^\tau = 0$ a.s. for all t. It suffices to show that $\mathbb{E}[Y_t^2] = 0$ for all t (for if $\{Y_t \neq 0\}$ has positive measure, then $\mathbb{E}[Y_t^2] > 0$).

Note that Y_t is a continuous martingale, and that $V_Y(t) \leq t$ for all t. Let $P = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ be a partition of $[0, t]$, and use the orthogonality of martingale increments to obtain

$$
\begin{aligned}
Y_t^2 &= \sum_{k=1}^{n} Y_{t_k}^2 - Y_{t_{k-1}}^2 \\
&= \sum_{k=1}^{n} (Y_{t_k} - Y_{t_{k-1}})^2 \\
&\leq V_Y(t) \max_k |Y_{t_k} - y_{t_{k-1}}|
\end{aligned}
$$

$$\leq K \max_k |Y_{t_k} - Y_{t_{k-1}}|$$

using the fact that $\sum_i a_i^2 \leq (\sum_i |a_i|) \cdot \max_i |a_i|$. Thus

$$\mathbb{E} Y_t^2 \leq K \mathbb{E}[\max_k |Y_{t_k} - Y_{t_{k-1}}|]$$

As $||P|| \to 0$, $\max_k |Y_{t_k} - Y_{t_{k-1}}| \downarrow 0$ a.s., by continuity of Y. By Dominated Convergence Theorem, because $\max_k |Y_{t_k} - Y_{t_{k-1}}| \leq 2K$, we see that $\mathbb{E}[\max_k |Y_{t_k} - Y_{t_{k-1}}|] \to 0$, as $||P|| \to 0$. so that $\mathbb{E}[Y_t^2] = 0$, as required.

We have now shown that $\mathbb{P}(Y_t = 0) = 1$ for all $t \geq 0$. It follows that

$$\mathbb{P}(Y_q = 0 \text{ for all } q \in \mathbb{Q}^+) = 1$$

Since Y_t is continuous, it now follows that

$$\mathbb{P}(Y_t = 0 \text{ for all } t \geq 0) = 1$$

$\dashv$

Since standard Brownian motion is a continuous martingale that is not constant, it cannot be of bounded variation. In general, therefore, the Ito integral $\int_0^T H_t \, dW_t$ cannot be interpreted as a Riemann–Stieltjes integral (unless H_t is itself of bounded variation, in which case we can use integration by parts).

By the way, this could also have been deduced from the fact, stated in the previous section (but not proved), that a function which is locally of bounded variation is necessarily differentiable almost everywhere.

3.5 Quadratic Variation and Covariation

The fact that the only continuous martingales of bounded variation are constant means that we will not be able to define stochastic integrals with respect to continous martingales in the Riemann–Stieltjes or Lebesgue–Stieltjes way. We need to replace the variation by something that is finite, and that is quadratic variation. Keep in mind that the series $\sum_n \frac{1}{n}$ diverges, but $\sum_n \frac{1}{n^2}$ converges. Essentially, we define the quadratic variation $\langle X \rangle_t$ of a process X_t to be

$$\langle X \rangle_t(\omega) = \lim_{||P|| \to 0} \sum_{k=1}^{n(P)} (X_{t_k}(\omega) - X_{t_{k-1}}(\omega))^2$$

where the limit is over successively finer partitions P. That's the intuitive idea, but we will need to be a bit more precise. The main aim of this section is to prove the following important result:

Theorem 3.5.1 *Let M_t be a continuous local martingale. Then there is a unique continuous increasing process Q_t with $Q_0 = 0$ such that*

$$M_t^2 - Q_t$$

is a continuous local martingale. Moreover, if $M \in c\mathcal{M}_0^2$, then $M_t^2 - Q_t$ is a UI martingale. The process Q_t is called the quadratic variation process *or* variance process *of M_t and is denoted by $\langle M \rangle_t$.*

$\dashv$

Please note the following easily verified identity (the "Summation By Parts" formula):

$$x_n y_n - x_0 y_0 = \sum_{k=1}^{n} x_{k-1}(y_k - y_{k-1}) + \sum_{k=1}^{n} y_{k-1}(x_k - x_{k-1}) + \sum_{k=1}^{n} (x_k - x_{k-1})(y_k - y_{k-1})$$

This formula is often applied with $y_k = x_k$:

$$x_n^2 - x_0^2 = 2\sum_{k=1}^{n} x_{k-1}(x_k - x_{k-1}) + \sum_{k=1}^{n} (x_k - x_{k-1})^2$$

Consider now a partition $\pi = \{0 = t_0 < t_1 < t_2 < \dots\}$ of the non–negative reals, with $t_n \uparrow \infty$. Define a function $k : \mathbb{R}^+ \to \mathbb{N}$ by $k(t) = \sup\{k : t_k < t\}$. Further define a process $Q_t^\pi(M)$ by

$$Q_t^\pi(M) = \sum_{t=0}^{k(t)} (M_{t_k} - M_{t_{k-1}})^2 + (M_t - M_{t_{k(t)}})^2$$

As the partition k gets finer and finer, $Q_t^\pi(M)(\omega)$ converges to the quadratic variation of the sample path $t \mapsto M_t(\omega)$ (provided it exists). Moreover, we have the following very nice property:

Proposition 3.5.2 *If M_t is a bounded continuous martingale, then so is $M_t^2 - Q_t^\pi(M)$.*

Proof: Let $Q_t = Q_t^\pi(M)$. It is clear that Q_t is continuous if M_t is, and thus $M_t^2 - Q_t$ is continuous. Note that if $r < s < t$, then

$$\begin{aligned}
\mathbb{E}[(M_t - M_r)^2|\mathcal{F}_s] &= \mathbb{E}[((M_t - M_s) + (M_s - M_r))^2|\mathcal{F}_s] \\
&= \mathbb{E}[(M_t - M_s)^2|\mathcal{F}_s] + (M_s - M_r)^2 \\
&= \mathbb{E}[M_t^2 - M_s^2|\mathcal{F}_s] + (M_s - M_r)^2
\end{aligned}$$

Now clearly

$$\begin{aligned}
Q_t - Q_s &= \sum_{k=k(s)+1}^{k(t)} (M_{t_k} - M_{t_{k-1}})^2 + (M_t - M_{t_{k(t)}})^2 - (M_s - M_{t_{k(s)}})^2 \\
&= (M_{t_{k(s)+1}} - M_{t_{k(s)}})^2 - (M_s - M_{t_{k(s)}})^2 + \sum_{k=k(s)+2}^{k(t)} (M_{t_k} - M_{t_{k-1}})^2 + (M_t - M_{t_{k(t)}})^2
\end{aligned}$$

Now take conditional expectations, using the formula at the beginning of the proof, with $r < s < t$ given by $r = t_{k(s)} < s < t = t_{k(s)+1}$:

$$\begin{aligned}\mathbb{E}[Q_t - Q_s|\mathcal{F}_s] &= \mathbb{E}[M^2_{t_{k(s)+1}} - M^2_s|\mathcal{F}_s] + (M_s - M_{t_{k(s)}})^2 \\ &\quad - (M_s - M_{t_{k(s)}})^2 + \sum_{k=k(s)+2}^{k(t)} \mathbb{E}[M^2_{t_k} - M^2_{t_{k-1}}|\mathcal{F}_s] + \mathbb{E}[M^2_t - M^2_{t_{k(t)}}|\mathcal{F}_s] \\ &= \mathbb{E}[M^2_t - M^2_s|\mathcal{F}_s]\end{aligned}$$

We now see that

$$\begin{aligned}\mathbb{E}[M^2_t - Q_t|\mathcal{F}_s] &= \mathbb{E}[M^2_t - (Q_t - Q_s)|\mathcal{F}_s] - Q_s \\ &= M^2_s - Q_s\end{aligned}$$

$\dashv$

Why does this not prove Theorem 3.5.1? The problem is that $Q^\pi_t(M)$ may not be an increasing function (in t). However, it is clear that if the partition π is finer than the partition Δ, then $Q^\pi_t(M) \geq Q^\Delta_t(M)$. In some texts, it is now shown that the $Q^\pi_t(M)$ converge in probability to some increasing continuous process (as the π's get finer), and this process is then called the quadratic variation. We shall take a different tack.

Given that $M \in \mathcal{M}^2_0$, define for each $n \in \mathbb{N}$ a sequence stopping times T^n_k by

$$T^n_0 = 0 \qquad T^n_{k+1} = \inf\{t > T^n_k : |M_t - M_{T^n_k}| > 2^{-n}\}$$

Further define $t^n_k = t \wedge T^n_k$. Put

$$Q^n_t = \sum_{k\geq 1}(M_{t^n_k} - M_{t^n_{k-1}})^2$$

We shall show that the processes A^n converge uniformly a.s. to an increasing continuous process Q_t and that $M^2_t - Q_t$ is a UI martingale. For this, we require the following two propositions:

Proposition 3.5.3 *Let M_t be a UI martingale, and let $\sigma \leq \tau$ be stopping times. Suppose that Z is a bounded $\mathcal{F}_\sigma$–measurable random variable. Define $N_t = Z(M_{\tau\wedge t} - M_{\sigma\wedge t})$. Then N_t is a UI martingale. Moreover, if $M_t \in \mathcal{M}^2_0$, then $N_t \in \mathcal{M}^2_0$ as well.*

Proof: We first deal with the UI case: By Theorem 2.2.14, it suffices to show that for every stopping time $\rho \leq \infty$ we have

$$\mathbb{E}|N_\rho| < \infty \qquad \mathbb{E}N_\rho = 0$$

If $c = \sup_\omega |Z(\omega)| < \infty$, then

$$\mathbb{E}|N_\rho| \leq c(\mathbb{E}[|M_{\tau\wedge\rho}| + |M_{\sigma\wedge\rho}|])$$

But $\mathbb{E}[|M_\infty||\mathcal{F}_{\tau\wedge\rho}] \geq |\mathbb{E}[M_\infty|\mathcal{F}_{\tau\wedge\rho}]| = |M_{\tau\wedge\rho}|$ by the Optional Sampling Theorem, so that $\mathbb{E}|M_{\tau\wedge\rho}| \leq \mathbb{E}|M_\infty| < \infty$. A similar argument shows that $\mathbb{E}|M_{\sigma\wedge\rho}| < \infty$ as well, and we conclude that $\mathbb{E}|N_\rho| < \infty$ for all stopping times ρ.

Next we must show that $\mathbb{E}N_\rho = 0$. For this, we apply the Monotone Class Theorem (Appendix A.2). Let $A \in \mathcal{F}_\sigma$, and first consider the case $Z = I_A$. Define random times τ_A, σ_A by

$$\tau_A(\omega) = \begin{cases} \tau(\omega) & \text{if } \omega \in A \\ +\infty & \text{else} \end{cases}$$

and similarly for σ_A. Then τ_A is a stopping time: Since $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$, we must have $A \cap \{\tau \leq t\} = \{\tau_A \leq t\} \in \mathcal{F}_t$ for all t. In the same way, it can be shown that σ_A is a stopping time. By the Optional Sampling Theorem, $\mathbb{E}[M_{\tau_A \wedge \rho}] = \mathbb{E}[M_{\sigma_A \wedge \rho}]$. Since $N_\rho = I_A(M_{\tau\wedge\rho} - M_{\sigma\wedge\rho}) = M_{\tau_A\wedge\rho} - M_{\sigma_A\wedge\rho}$, the result follows for indicator Z.

Now let $\mathcal{H}_\rho = \{Z : Z \text{ is } \mathcal{F}_\sigma\text{–measurable}, \mathbb{E}Z(M_{\tau\wedge\rho} - M_{\sigma\wedge\rho}) = 0\}$. Then, invoking the Bounded Convergence Theorem, $\mathcal{H}_\rho$ satisfies the hypotheses of the Monotone Class Theorem, and $I_A \in \mathcal{H}_\rho$ for each $A \in \mathcal{F}_\sigma$. Hence every bounded $\mathcal{F}_\sigma$–measurable random variable belongs to $\mathcal{H}_\rho$.

Next suppose that $M_t \in \mathcal{M}_0^2$. From the aforegoing, it is obvious that N_t is a UI martingale, null at zero. Moreover,

$$\begin{aligned}\mathbb{E}[N_t^2] &\leq c^2\mathbb{E}[(M_{\tau\wedge t} - M_{\sigma\wedge t})^2] \\ &= c^2\mathbb{E}[\mathbb{E}[M_{\tau\wedge t}^2 - 2M_{\tau\wedge t}M_{\sigma\wedge t} + M_{\sigma\wedge t}^2|\mathcal{F}_{\sigma\wedge t}]] \\ &= c^2\mathbb{E}[M_{\tau\wedge t}^2 - M_{\sigma\wedge t}^2]\end{aligned}$$

by the Optional Sampling Theorem applied to the UI martingale M_t. Hence $\mathbb{E}[N_t^2] \leq c^2\mathbb{E}[M_\infty^2] < \infty$, i.e. $N_t \in \mathcal{M}_0^2$.

$\dashv$

Proposition 3.5.4 *Let $M \in \mathcal{M}_0^2$, let $0 \leq \tau_0 \leq \tau_1 \leq \dots$ be stopping times, and let Z_k be $\mathcal{F}_{\tau_k}$–measurable random variables with $|Z_k| \leq c$ for each k. Let N be a bounded process of the form*

$$N_t = \sum_{k\geq 1} Z_{k-1}(M_{\tau_k\wedge t} - M_{\tau_{k-1}\wedge t})$$

Then $N \in \mathcal{M}_0^2$, and $||N||_2 \leq c||M||_2$.

Proof: Consider first the following processes, with finitely many terms each:

$$N_t^n = \sum_{k=1}^n Z_{k-1}(M_{\tau_k\wedge t} - M_{\tau_{k-1}\wedge t})$$

By the previous proposition, each term $Z_{k-1}(M_{\tau\wedge t} - M_{\tau_{k-1}\wedge t})$ belongs to $\mathcal{M}_0^2$, and hence so does each N_t^n (because $\mathcal{M}_0^2$ is a vector space). Indeed

$$\begin{aligned}\mathbb{E}[(N_t^n)^2] &\leq \mathbb{E}[\sum_{k=1}^n Z_{k-1}(M_{\tau_k\wedge t} - M_{\tau_{k-1}\wedge t})^2] \\ &\leq c^2\sum_{k=1}^n \mathbb{E}[(M_{\tau_k\wedge t} - M_{\tau_{k-1}\wedge t})^2] \\ &= c^2\sum_{k=1}^n \mathbb{E}[M_{\tau_k\wedge t}^2 - M_{\tau_{k-1}\wedge t}^2] \\ &\leq c^2\mathbb{E}[M_\infty^2]\end{aligned}$$

so that $||N^n||_2 \leq c||M||_2$.

Now fix t. Clearly $N_t^n \to N_t$ a.s. as $n \to \infty$. By the aforegoing, $\{N_t^n : n \in \mathbb{N}\}$ is bounded in $\mathcal{L}^2$, and thus UI. We thus have $N_t^n \to N_t$ in $\mathcal{L}^1$ as well (by Theorem 1.6.8). Thus $\mathbb{E}[N_t^n|\mathcal{F}_s] \to \mathbb{E}[N_t|\mathcal{F}_s]$ in $\mathcal{L}^1$ for $s \leq t$. However, also $\mathbb{E}[N_t^n|\mathcal{F}_s] = N_s^n \to N_s$ in $\mathcal{L}^1$. It follows that $\mathbb{E}[N_t|\mathcal{F}_s] = N_s$ a.s., i.e. that N_t is a martingale. By Fatou's Lemma,

$$\mathbb{E}[N_t^2] \leq \uparrow \lim_n \mathbb{E}[(N_t^n)^2] \leq c^2 \mathbb{E}[M_\infty^2]$$

which shows that $N \in \mathcal{M}_0^2$ with $||N||_2 \leq c||M||_2$.

$\dashv$

Proof of Theorem 3.3.1: First assume that M is a bounded martingale null at zero. Proving *uniqueness* is easy: If $M_t^2 - Q_t$ and $M_t^2 - Q_t'$ are both continuous martingales, where Q_t, Q_t' are increasing continuous processes null at zero, then the difference

$$(M_t^2 - Q_t') - (M_t^2 - Q_t) = Q_t - Q_t'$$

is a martingale, and, moreover, locally of bounded variation (being a difference of two increasing processes). Thus $Q_t - Q_t'$ is constant, and since $Q_0 = Q_0' = 0$, we have $Q_t = Q_t'$.

Next we worry about *existence*: Define for each $n \in \mathbb{N}$ a sequence stopping times T_k^n by

$$T_0^n = 0 \qquad T_{k+1}^n = \inf\{t > T_k^n : |M_t - M_{T_k^n}| > 2^{-n}\}$$

Further define $t_k^n = t \wedge T_k^n$. Put

$$Q_t^n = \sum_{k \geq 1} (M_{t_k^n} - M_{t_{k-1}^n})^2$$

By the summation by parts formula

$$\begin{aligned} M_t^2 &= 2\sum_{k\geq 1} M_{t_{k-1}^n}(M_{t_k^n} - M_{t_{k-1}^n}) + \sum_{k \geq 1}(M_{t_k^n} - M_{t_{k-1}^n})^2 \\ &= 2N_t^n + Q_t^n \end{aligned}$$

where the sum in the first term defines N_t^n. Note that by the previous proposition, $N^n \in \mathcal{M}_0^2$ for each n, and thus $M_t^2 - Q_t^n$ is a martingale, for each n.

We shall now show that the processes A^n converge uniformly a.s. Note that the t_k^n form successively finer partitions of $(0, t]$ as n increases, and thus that each t_j^{n-1} is equal to some t_k^n. Define

$$s_k^n = \max\{t_j^{n-1} : t_j^{n-1} \leq t_k^n\}$$

so that s_k^n is the biggest t_j^{n-1} which lies below t_k^n. Note that the s_k^n are not necessarily distinct.

Now fix j and consider the term $M_{t_{j-1}^n}(M_{t_j^n} - M_{t_{j-1}^n})$. Also consider the set

$$\{t_k^{n+1} : s_k^{n+1} = t_{j-1}^n\} = \{t_m^{n+1}, t_{m+1}^{n+1}, \ldots, t_{m+l-1}^{n+1}\}$$

i.e. the set of all t_k^{n+1} such that $t_{j-1}^n \leq t_k^{n+1} < t_j^n$. Note that $t_{m+l}^{n+1} = t_j^n$. Then $M_{t_{j-1}^n} = M_{s_m^{n+1}} = M_{s_{m+1}^{n+1}} = \cdots = M_{s_{m+l}^{n+1}}$, and so

$$M_{t_{j-1}^n}(M_{t_j^n} - M_{t_{j-1}^n}) = \sum_{k=m+1}^{m+l} M_{s_{k-1}^{n+1}}(M_{t_k^{n+1}} - M_{t_{k-1}^{n+1}})$$

It follows that

$$N_t^n = \sum_{j\geq 1} M_{t_{j-1}^n}(M_{t_j^n} - M_{t_{j-1}^n}) = \sum_{k\geq 1} M_{s_{k-1}^{n+1}}(M_{t_k^{n+1}} - M_{t_{k-1}^{n+1}})$$

and thus

$$N_t^{n+1} - N_t^n = \sum_{k\geq 1}(M_{t_{k-1}^{n+1}} - M_{s_{k-1}^{n+1}})(M_{t_k^{n+1}} - M_{t_{k-1}^{n+1}})$$

which is of the form

$$N_t^{n+1} - N_t^n = \sum_{k\geq 1} Z_{k-1}(M_{t_k^{n+1}} - M_{t_{k-1}^{n+1}})$$

where each $Z_{k-1} = (M_{t_{k-1}^{n+1}} - M_{s_{k-1}^{n+1}})$ is $\mathcal{F}_{t_{k-1}^{n+1}}$–measurable. Now clearly each $|Z_{k-1}| \leq 2 \cdot 2^{-(n+1)} = 2^{-n}$, and thus by Proposition 3.3.4, each $N^{n+1} - N^n$ is a martingale in $c\mathcal{M}_0^2$ with

$$||N^{n+1} - N^n||_2 \leq 2^{-n}||M||_2$$

It follows that the sequence N^n is a Cauchy sequence in $c\mathcal{M}_0^2$, and thus it converges a.s. uniformly to some martingale $N \in c\mathcal{M}_0^2$. It follows that the processes Q_t^n converges a.s. uniformly to some process Q_t. Then $M_t^2 = 2N_t + Q_t$. Also, Q_t is continuous, being the uniform limit of the continuous Q^n.

Now it is clear that even though the process Q^n may not be increasing, we certainly have $Q^n_{T_k^n} \leq Q^n_{T_{k+1}^n}$ for each k. By definition of Q we thus see that $Q_{T_k^n} \leq Q_{T_{k+1}^n}$ (for each n and k). For each $\omega \in \Omega$, define a countable subset $J(\omega)$ of $\mathbb{R}^+$ by

$$J_n(\omega) = \{T_k^n(\omega) : k \in \mathbb{N}\} \qquad J(\omega) = \bigcup_n J_n(\omega)$$

Then each $A^n(\omega)$ is increasing on $J_n(\omega)$, and thus A is increasing on $J(\omega)$ (and thus on the closure of $J(\omega)$. Now suppose that the interval I is disjoint from $J(\omega)$. Then no T_k^n belongs to I. Since $M(\omega)$ is continuous (a.s.) it follows that $M(\omega)$ is constant on I. But then each $Q^n(\omega)$ is constant on I as well, and hence $Q(\omega)$ is constant on I. Since "constant" is a special case of "increasing", we see that Q is (a.s.) increasing.

We have now proved that quadratic variation exists for bounded continous martingales. To deal with the general case, let M be an arbitrary continuous local martingale (null at zero), and choose a localizing sequence of stopping times τ_n such that each stopped process M^{τ_n} is a bounded martingale. By what we have just proved, there exists a unique increasing continuous process Q^n such that $M^2_{\tau_n \wedge t} - Q_t^n$ is a martingale for each n. However, uniqueness of Q^n ensures that $Q_t^{n+1}(\omega) = Q_t^n(\omega)$ whenever $t < \tau_n$. Thus we can define $Q_t(\omega)$ by

$$Q_t(\omega) = Q_t^n(\omega) \quad \text{where } n \text{ is such that } t \leq \tau_n(\omega)$$

(Such n exists because $\tau_n \uparrow \infty$.) Then Q is clearly continuous and increasing. Moreover, the stopped process

$$(M_t^2 - Q_t)^{\tau_n} = M^2_{\tau_n \wedge t} - Q_{\tau_n \wedge t}$$

is a martingale by construction, and thus $M_t^2 - Q_t$ is a local martingale.

$\dashv$

Different texts will give different definitions of quadratic variation. The advantage of the above construction is that it is defined pathwise, and thus we will be able to prove several a.s. results, whereas many of the other definitions will only yield results that are true in probability. One immediate consequence of the pathwise definition is that the quadratic variation does not depend on the probability measure. Implicitly, most of our definitions are a.s., so if $\mathbb{P}, \mathbb{Q}$ are equivalent probability measures, then the quadratic variation of a stochastic process X under $\mathbb{P}$ is the same as its quadratic variation under $\mathbb{Q}$.

Chapter 4

Outline of Stochastic Integration

4.1 L^2–Theory of the Stochastic Integral

Throughout, let $M, N \in c\mathcal{M}_0^2$ be continuous square–integrable martingales starting at 0.

4.1.1 Basic Integrands

A *basic predictable process* is analogous to a buy–and–hold strategy $H := CI_{(t_1,t_2]}$: Buy C–many shares M at time t_1, and sell them at t_2. The amount bought at time t_1 must be known at time t_1, i.e. C is a bounded $\mathcal{F}_{t_1}$–measurable RV. The gain G_t at time t is then

$$G_t := \begin{cases} 0 & \text{if } t < t_1 \\ C(M_t - M_{t_1}) & \text{if } t_1 \leq t \leq t_2 \\ C(M_{t_2} - M_{t_1}) & \text{if } t_2 < t \end{cases} \qquad \text{i.e.} \qquad G_t := C(M_{t_2\wedge t} - M_{t_1\wedge t})$$

We define

$$(H \bullet M)_t \equiv \int_0^t H_s \, dM_s := C(M_{t_2\wedge t} - M_{t_1\wedge t})$$

Proposition 4.1.1 $H \bullet M \in c\mathcal{M}_0^2$.

Proof: The gist: Let τ be a stopping time. Then $\mathbb{E}[(H \bullet M)_\tau] = \mathbb{E}[C(M_{t_2\wedge\tau} - M_{t_1\wedge\tau})] = \mathbb{E}[C\mathbb{E}[M^\tau_{t_2} - M^\tau_{t_1}|\mathcal{F}_{t_1}]] = 0$, as M^τ is a martingale. It follows that $(H \bullet M)$ is a UI martingale.

Furthermore, by orthogonality of martingale increments, $\mathbb{E}[(H\bullet)^2_\infty] = \mathbb{E}[C^2(M^2_{t_2} - M^2_{t_1})] = \mathbb{E}[C^2\mathbb{E}[(M_{t_2} - M_{t_1})|\mathcal{F}_{t_1}]] = \mathbb{E}[C^2\mathbb{E}[M^2_{t_2} - M^2_{t_1}|\mathcal{F}_{t_1}]] = \mathbb{E}[C^2(M^2_{t_2} - M^2_{t_1})]] < \infty$ as C is bounded. Hence $H \bullet M$ is square–integrable.

$\dashv$

4.1.2 Simple Integrands

Now consider a more dynamic trading strategy H. Trade at dates $0 \leq t_0 < t_1 < \cdots < t_n$ only, to get

$$H := \sum_{k=1}^{n} C_{k-1} I_{(t_{k-1},t_k]} \qquad \text{where } C_{k-1} \in \mathcal{F}_{t_{k-1}} \text{ is bounded}$$

Such a linear combination of basic integrands is called a *simple predictable process.* Extend the integral from basic to simple integrands by demanding that linearity be preserved, i.e. define

$$(H\bullet M)_t \equiv \int_0^t H_s \, dM_s := \sum_{k=1}^{n} C_{k-1}(M_{t_k\wedge t} - M_{t_{k-1}\wedge t}) = \sum_{k=1}^{k(t)} C_{k-1}(M_{t_k} - M_{t_{k-1}}) + C_{k(t)}(M_t - M_{t_{k(t)}})$$

wheret $k(t) := \max\{k : t_k \leq t\}$

Proposition 4.1.2 $H \bullet M \in c\mathcal{M}_0^2$.

Proof: The gist: By Propn. 4.1.1, each $C_{k-1}(M_{t_k\wedge t} - M_{t_{k-1}\wedge t})$ is a martingale in $c\mathcal{M}_0^2$. Since $H \bullet M$ is a sum of these martingales, it too is $\in c\mathcal{M}_0^2$.

$\dashv$

Observe that the set $\mathcal{S}$ of simple predictable processes is a vector space. Furthermore, the integral is a linera operator, i.e. if H, K are simple predictable, and $\alpha, \beta \in \mathbb{R}$, then

$$(\alpha H + \beta K) \bullet M = \alpha(H \bullet M) + \beta(K \bullet M)$$

i.e.

$$\int_0^t \alpha H_s + \beta K_s \, dM_s = \alpha \int_0^t H_s \, dM_s + \beta \int_0^t K_s \, dM_s$$

Proposition 4.1.3 *If H is a simple predictable process, then*

$$||H \bullet M||^2_{\mathcal{M}^2} = \mathbb{E}\left[\int_0^\infty H_s^2 \, d[M]_s\right]$$

Comment: The integral on the right is (for each outcome $\omega \in \Omega$) an ordinary Riemann–Stieltjes (or Lebesgue–Stieltjes) integral, as the process $[M]_t$ is increasing, and thus of locally finite variation.

Proof: $||H \bullet M||_{\mathcal{M}^2} := \mathbb{E}[(H \bullet M)^2_\infty] = \mathbb{E}\Big(\sum_{k=1}^n C_{k-1}(M_{t_k} - M_{t_{k-1}})\Big)^2$. Now

$$\Big(\sum_{k=1}^{n} C_{k-1}(M_{t_k} - M_{t_{k-1}})\Big)^2 = \sum_{k=1}^{n} C_{k-1}^2(M_{t_k} - M_{t_{k-1}})^2 + 2\sum_{k=1}^{n}\sum_{j<k} C_{k-1}C_{j-1}(M_{t_k} - M_{t_{k-1}})(M_{t_j} - M_{t_{j-1}})$$

By orthogonality of martingale increments and the tower property,

$$\mathbb{E}\left[\sum_{k=1}^{n} C_{k-1}^2(M_{t_k} - M_{t_{k-1}})^2\right] + 2\mathbb{E}\left[\sum_{k=1}^{n}\sum_{j<k} C_{k-1}C_{j-1}(M_{t_k} - M_{t_{k-1}})(M_{t_j} - M_{t_{j-1}})\right]$$

$$= \mathbb{E}\left[\sum_{k=1}^{n} C_{k-1}^2(M_{t_k}^2 - M_{t_{k-1}}^2)\right] + 0$$

Now $M_t^2 - [M]_t$ is a UI martingale, and hence $\mathbb{E}[M_{t_k}^2 - M_{t_{k-1}}^2 | \mathcal{F}_{t_{k-1}}] = \mathbb{E}[[M]_{t_k} - [M]_{t_{k-1}} | \mathcal{F}_{t_{k-1}}]$. It follows that

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^n C_{k-1}^2 (M_{t_k}^2 - M_{t_{k-1}}^2)\right] &= \mathbb{E}\left[\sum_{k=1}^n C_{k-1}^2 ([M]_{t_k} - [M]_{t_{k-1}})\right] \\
&= \mathbb{E}\left[\sum_{k=1}^n \int_0^\infty C_{k-1}^2 I_{(t_{k-1}, t_k]} \, d[M]_s\right] \\
&= \mathbb{E}\left[\int_0^\infty H_t^2 \, d[M]_s\right]
\end{aligned}$$

$\dashv$

4.1.3 The space $L^2(M)$

Given an $M \in c\mathcal{M}_0^2$, define a function $||\cdot||_{L^2(M)}$ on the set of predictable processes by

$$||H||_{L^2(M)} := \left(\mathbb{E}\left[\int_0^\infty H_s^2 \, d[M]_s\right]\right)^{\frac{1}{2}}$$

This looks a bit like an L^2–norm.

Let
$$L^2(M) := \Big\{\text{set of all predictable processes } H \text{ with } ||H||_{L^2(M)} < \infty\Big\}$$
(where two processes are regarded as equal if they are indistinguishable).

Here are some facts:

- $||\cdot||_{L^2(M)}$ makes $L^2(M)$ into a normed vector space.
- Every simple predictable process belongs to $L^2(M)$.
- The set $\mathcal{S}$ of simple predictable processes is dense in $L^2(M)$, i.e. if $H \in L^2(M)$ then there is a sequence $(H^{(n)})_n$ of simple predictable processes such that $H^{(n)} \to H$ in $L^2(M)$, i.e. $||H^{(n)} - H||_{L^2(M)} \to 0$ as $n \to \infty$.

Further observe that we have the following *isometry*:

Corollary 4.1.4 (Itô Isometry) *If $H \in \mathcal{S}$ is a simple predictable process, then*

$$||H||_{L^2(M)} = ||H \cdot M||_{\mathcal{M}^2}$$

4.1.4 The Stochastic Integral on $L^2(M)$

We now use the Itô isometry and the fact that $\mathcal{S}$ is dense in $L^2(M)$ to lift the stochastic integral from $\mathcal{S}$ to all of $L^2(M)$. The basic idea is as follows: Let $H \in L^2(M)$. We want to define $H \bullet M$.

- Choose a sequence of simple predictable processes $H^{(n)} \in \mathcal{S}$ so that $||H^{(n)} - H||_{L^2(M)} \to 0$.

- Then the sequence $H^{(n)}$ is a Cauchy sequence in $L^2(M)$ (since any convergent sequence is a Cauchy sequence), i.e.
$$||H^{(n)} - H^{(m)}||_{L^2(M)} \to 0 \qquad \text{as} \qquad n, m \to \infty$$
- Now the stochastic integral has already been defined for simple predictable processes, i.e. $H^{(n)} \bullet M$ has already been defined, for each n, and each $H^{(n)} \bullet M \in c\mathcal{M}_0^2$.
- Furthermore, $||H^{(n)} \bullet M - H^{(m)} \bullet M||_{\mathcal{M}^2} = ||(H^{(n)} - H^{(m)}) \bullet M||_{\mathcal{M}^2} = ||H^{(n)} - H^{(m)}||_{L^2(M)}$. Thus $(H^{(n)} \bullet M)_n$ is a Cauchy sequence $c\mathcal{M}_0^2$.
- But $c\mathcal{M}_0^2$ is a Hilbert space, and hence complete. Thus the Cauchy sequence $(H^{(n)} \bullet M)_n$ converges to some martingale $N \in c\mathcal{M}_0^2$.
- Now we define $H \bullet M$ to be that limit N.
- Automatically we have $H \bullet M \in c\mathcal{M}_0^2$.

Proposition 4.1.5 (Itô Isometry *If $H \in L^2(M)$, then $||H||_{L^2(M)} = ||H \bullet M||_{\mathcal{M}^2}$.*

Proof: Choose $H^{(n)} \in \mathcal{S}$ so that $H^{(n)} \to H$ in $L^2(M)$. It follows that $||H^{(n)}||_{L^2(M)} \to ||H||_{L^2(M)}$[1]. Similarly, since $H^{(n)} \bullet M \to H \bullet M$ in $c\mathcal{M}_0^2$, we have $||H^{(n)} \bullet M||_{\mathcal{M}^2} \to ||H \bullet M||_{\mathcal{M}^2}$. The Itô isometry for simple predictable processes yields that $||H^{(n)}||_{L^2(M)} = ||H^{(n)} \bullet M||_{\mathcal{M}^2}$ for each n. Taking limits, we see that
$$|||H||_{L^2(M)} = \lim_n ||H^{(n)}||_{L^2(M)} = \lim_n ||H^{(n)} \bullet M||_{\mathcal{M}^2} = ||H \bullet M||_{\mathcal{M}^2}$$
$\dashv$

A similar argument shows that the stochastic integral is linear.

The following fact is sometimes useful, and easy to believe. It says that stopping the stochastic integral $(H \bullet M)$ at a stopping time τ is equivalent to either stopping the martingale M at time τ (so that $dM_t^\tau = 0$ for $t \geq \tau$), or setting the integrand H to zero after time τ.

Proposition 4.1.6 *If τ is a stopping time, then*
$$(H \bullet M)^\tau = H \bullet M^\tau = HI_{(0,\tau]} \bullet M$$
i.e.
$$\int_0^{\tau \wedge t} H_s \, dM_s = \int_0^t H_s \, dM_s^\tau = \int_0^t H_s I_{(0,\tau]} \, dM_s$$

Proof: We will prove that $(H \bullet M)^\tau = H \bullet M^\tau$: If H is a simple predictable process, $H := \sum_k C_{k-1} I_{(t_{k-1}, t_k]}$, then
$$\begin{aligned}
(H \bullet M)_t^\tau &= \Big(\sum_k C_{k-1}(M_{t_k \wedge t} - M_{t_{k-1} \wedge t}\Big)^\tau \\
&= \sum_k C_{k-1}(M_{t_k \wedge t \wedge \tau} - M_{t_{k-1} \wedge t \wedge \tau}) \\
&= \sum_k C_{k-1}(M^\tau_{t_k \wedge t} - M^\tau_{t_{k-1} \wedge t}) \\
&= (H \bullet M^\tau)_t
\end{aligned}$$

[1]Observe that if $x_n \to x$ in a normed space, then $|x_n - x|| \to 0$, and since $0 \leq \Big| ||x_n|| - ||x|| \Big| \leq ||x_n - x||$, we have $||x_n|| \to ||x||$.

which proves the result if H is simple.

Now let $H \in L^2(M)$. Choose a sequence $H^{(n)}$ of simple predictable processes such that $||H^{(n)} - H||_{L^2(M)} \to 0$. Then also $||H^{(n)} \bullet M - H \bullet M||_{\mathcal{M}^2} \to 0$. Now by the Optional Sampling Theorem and Jensen's Inequality,

$$\begin{aligned}
||(H^{(n)} \bullet M)^\tau - (H \bullet M)^\tau||^2_{\mathcal{M}^2} &= \mathbb{E}\left[\Big((H \bullet M)_\tau - (H \bullet M)_\tau\Big)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}[(H^{(n)} \bullet M)_\infty - (H \bullet M)_\infty | \mathcal{F}_\tau]^2\right] \\
&\leq \mathbb{E}\left[\Big((H^{(n)} \bullet M)_\infty - (H \bullet M)_\infty\Big)^2\right] \\
&= ||H^{(n)} \bullet M - H \bullet M||_{\mathcal{M}^2} \to 0
\end{aligned}$$

Thus $(H^{(n)} \bullet M)^\tau \to (H \bullet M)^\tau$ in $c\mathcal{M}_0^2$. However, $(H^{(n)} \bullet M)^\tau = (H^{(n)} \bullet M^\tau)$, so also $(H^{(n)} \bullet M)^\tau \to (H \bullet M^\tau)$. Uniqueness of limits shows $(H \bullet M)^\tau = (H \bullet M^\tau)$, as required.

$\dashv$

4.1.5 An Example

This example contains some heuristic calculations, and is not watertight. Consider the stochastic integral $\int_0^t B_s \; dB_s$. Here the integrand is $H = BI(0,t]$, and the integrator M is B. Recall that $[B]_t = t$.

Note that Brownian motion is *not* a square–integrable martingale, since $\sup_{s \geq 0} \mathbb{E}[B_s^2] = \infty$. However, if $t \geq 0$ is fixed, then the stopped Brownian motion B^t (defined by $B^t_s := B_{s \wedge t}$ is a square–integrable martingale, as $\sup_{s \geq 0} \mathbb{E}[(B^t_s)^2] = \mathbb{E}[B^2_t] = t < \infty$. Then $\int_0^t H_s \; dB_s = \int_0^\infty H_s \; dB^t_s$ as $dB^t_s = 0$ for $s \geq t$. Hence the theory developed so far applies.

Let $P \equiv 0 = t_0 < t_1 < t_2 < \cdots < t_{n(P)} = t$ be a partition of $[0,t]$, and define simple predictable processes

$$H^{(P)} := \sum_{k=1}^{n(P)} B_{t_{k-1}} I_{(t_{k-1}, t_k]}$$

Recall that the mesh $\sigma(P)$ of the partition P is defined by $\sigma(P) := \max\{|t_k - t_{k-1}| : k = 1, \ldots, n(P)\}$.

Now

$$\begin{aligned}
||H^{(P)} - H||^2_{L^2(M)} &= \mathbb{E}\left[\sum_{k=1}^{n(P)} \int_{t_{k-1}}^{t_k} (B_{t_{k-1}} - B_t)^2 \; dt\right] \\
&= \sum_{k=1}^{n(P)} \int_{t_{k-1}}^{t_k} (t - t_{k-1}) \; dt \\
&= \tfrac{1}{2} \sum_{k=1}^{n(P)} (t_k - t_{k-1})^2
\end{aligned}$$

Now as $\sigma(P) \to 0$, we have $\sum_{k=1}^{n(P)} (t_k - t_{k-1})^2 \to [t]$, the quadratic variation of the function $f(t) = t$, which is zero. Hence $H^{(P)} \to H$ in $L^2(M)$ as $\sigma(P) \to 0$, and hence $H^{(P)} \bullet B \to H \bullet B$ in $c\mathcal{M}_0^2$. It follows that $\sum_{k=1}^{n(P)} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) \to \int_0^t H_t \; dB_t$.

Now note that $b(a-b) = \frac{1}{2}(a+b)(a-b) - \frac{1}{2}(a-b)^2 = \frac{1}{2}(a^2-b^2) - \frac{1}{2}(a-b)^2$. With $a := B_{t_k}$ and $b := B_{t_{k-1}}$, we get

$$\sum_{k=1}^{n(P)} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) = \frac{1}{2}\underbrace{\sum_{k=1}^{n(P)}(B_{t_k}^2 - B_{t_{k-1}}^2)}_{(I)} - \frac{1}{2}\underbrace{\sum_{k=1}^{n(P)}(B_{t_k} - B_{t_{k-1}})^2}_{(II)}$$

The sum (I) is telescoping, and sums to $B_{t_{n(P)}}^2 - B_{t_0}^2 = B_t^2 - B_0^2 = B_t^2$.

The sum (II) converges (in probability) to the quadratic variation $[B]_t = t$ (as $\sigma(P) \to 0$).

Hence $\sum_{k=1}^{n(P)} B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})$ converges (in probability) to $\frac{1}{2}(B_t^2 - t)$. Since it also converges to $\int_0^t B_s \, dB_s$ (in L^2, and hence in probability), we see that

$$\int_0^t B_t \, dB_t = \frac{1}{2}(B_t^2 - t)$$

We already knew that $B_t^2 - t$ is a martingale.

4.1.6 Approximation

Suppose that H is a left–continuous adapted process, and that $P \equiv 0 = t_0 < t_1 < \cdots < t_{n(P)} = t$ is a partition of $[0,t]$. For $s \in [0,t]$, ket $k(s) := \max\{k : t_k \le s\}$. Then as $\sigma(P) \to 0$, we have $t_{k(s)} \uparrow s$. Since H is left–continuous, we see that $H_{t_{k(s)}} \to H_s$ a.s. Thus if we define the simple predictable process $H^{(P)}$ by

$$H^{(P)} = \sum_{k=1}^{n(P)} H_{t_{k-1}} I_{(t_{k-1}, t_k]}$$

then $H_s^{(P)} \to H_s$ a.s. as $\sigma(P) \to 0$, for all $s \in [0,t]$. Thus the simple predictable processes $H^{(P)}$ form better and better approximations to H.

Now note that $\int_0^t H_s^{(P)} \, dM_s = \sum_{k=1}^{n(P)} H_{t_{k-1}}(M_{t_k} - M_{t_{k-1}})$

The following fact is therefore hopefully not too hard to believe:

Proposition 4.1.7 *If $H \in L^2(M)$ is left–continuous, and $P \equiv 0 = t_0 < t_1 < \cdots < t_{n(P)} = t$ is a partition of $[0,t]$. Then*

$$\sum_{k=1}^{n(P)} H_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) \to \int_0^t H_s \, dM_s$$

as $\sigma(P) \to 0$.

This represents the stochastic integral as a lmit of left–hand Riemann–Stieltjes sums.

We now have two approximations that we will use repeatedly. Given partitions: $P \equiv 0 = t_0 < t_1 < \cdots < t_{n(P)} = t$ of $[0,t]$, with $\sigma(P) \to 0$.

1. The quadratic covariation $[M, N]$ of two martingales can be approximated by sums:

$$[M, N]_t \approx \sum_{k=1}^{n(P)} (M_{t_k} - M_{t_{k-1}})(N_{t_k} - N_{t_{k-1}}) = \sum_k \Delta_k M \cdot \Delta_k N \quad \text{as } \sigma(P) \to 0$$

   In particular, when $M = N$, we have $[M]_t = [M, M]_t \approx \sum_k (\Delta_k M)^2$.

   We thus have:

$$\Delta_k[M, N] \approx \Delta_k M \cdot \Delta_k N \qquad \Delta_k[M] \approx (\Delta_k M)^2$$

2. The stochastic integral can be approximated by left–hand Riemann–Stieltjes sums

$$\int_0^t H_s \, dM_s \approx \sum_{k=1}^{n(P)} H_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) = \sum_k H_{t_{k-1}} \cdot \Delta_k M \quad \text{as } \sigma(P) \to 0$$

   We thus have:

$$\Delta_k(H \bullet M) \approx H_{t_{k-1}} \Delta_k M$$

4.1.7 Quadratic Variation and Covariation of Stochastic Integrals

What is the quadratic variation of $H \bullet M$?

This (and more) is answered by the following proposition:

Proposition 4.1.8 *Let $M, N \in c\mathcal{M}_0^2$, and $H \in L^(M), K \in L^2(N)$. Then*

$$[H \bullet M, K \bullet N]_t = \int_0^t H_s K_s \, d[M, N]_s$$

Proof: Some rather complicated analysis shows that the following approximations are OK:

$$\begin{aligned}
[H \bullet M, K \bullet N]_t &\approx \sum_k \Delta_k(H \bullet M) \cdot \Delta_k(K \bullet N) \\
&\approx \sum_k H_{t_{k-1}} K_{t_{k-1}} \Delta_k M \Delta_k N \qquad \text{since } \Delta_k(H \bullet M) \approx H_{t_{k-1}} \Delta_k M \\
&\approx \sum_k H_{t_{k-1}} K_{t_{k-1}} \, \Delta_k[M, N] \qquad \text{since } \Delta_k M \cdot \Delta_k N \approx \Delta_k[M, N] \\
&\approx \int_0^t H_s K_s \, d[M, N]_s
\end{aligned}$$

⊣

4.1.8 The Associative Law

Suppose that $M \in c\mathcal{M}_0^2$ and that $H \in L^2(M)$. Since $H \bullet M$ is again a martingale in $c\mathcal{M}_0^2$, we may integrate with respect to it, i.e. given another predictable process K, we may ask:

What is $K \bullet (H \bullet M)$?

Proposition 4.1.9

$$K \bullet (H \bullet M) = (KH) \bullet M$$

Proof: Some more complicated analysis shows that the following approximations are OK:

$$\begin{aligned}
(K \bullet (H \bullet M))_t &\approx \sum_k K_{t_{k-1}} \Delta_k (H \bullet M) && \text{Lefthand Riemann–Stieltjes sum} \\
&\approx \sum_k K_{t_{k-1}} H_{t_{k-1}} \Delta_k M && \text{since } \Delta_k (H \bullet M) \approx H_{t_{k-1}} \Delta_k M \\
&\approx \int_0^t K_s H_s \, dM_s
\end{aligned}$$

⊣

4.2 Integration w.r.t. Semimartingales

4.2.1 Continuous Local Martingales

A stochastic process M is said to be a *continuous local martingale* if and only if there is a sequence of stopping times $\tau_n \uparrow \infty$ a.s. such that each stopped process M^{τ_n} is a continuous martingale. Such a sequence of stopping times is called a *localizing sequence* for M.

If τ_n is a localizing sequence for M, then $\tau_n \wedge t \to t$ as $n \to \infty$, and hence $M_t^{\tau_n} := M_{\tau_n \wedge t} \to M_t$ a.s. Observe that if $\tau_n \leq \tau_m$ then $M_t^{\tau_n} = M_t^{\tau_m}$ for $t \leq \tau_n$. By choosing τ_n intelligently, we can ensure that the stopped processes M^{τ_n} are "nice". Then, by taking limits $N \to \infty$, , we can lift results from"nice" martingales to continuous local martingales. This process is called *localization.*

Observe:

- Since stopped martingales are martingales, every martingale is a local martingale. The converse is not true.
- By choosing τ_n intelligently (e.g. $\tau_n := \inf\{t : |M_t| = n\}$, we can ensure that each M^{τ_n} is a bounded martingale, and hence square–integrable.
- If each M^{τ_n} is a square–integrable continuous martingale, then the quadratic variations $[M^{\tau_n}]_t$ and stochastic integrals $H \bullet M^{\tau_n}$ have already been defined.
- If each $M^{\tau_n} \in c\mathcal{M}_0^2$, then each M^{τ_n} has a quadratic variation process $[M^{\tau_n}]_t$. We may then define the quadratic variation $[M]$ of the continuous local martingale M by taking limits: $[M]_t = \lim_n [M^{\tau_n}]_t$. Since $M_t^{\tau_n} = M_t^{\tau_m}$ when $t \leq \tau_n \leq \tau_m$, it is easy to see that $[M^{\tau_n}]_t = [M]_t^{\tau_n}$.
- Now $(M_t^2 - [M]_t)^{\tau_n} = (M_t^{\tau_n})^2 - [M]_t^{\tau_n} = (M_t^{\tau_n})^2 = [M^{\tau_n}]_t$ is a continuous martingale. Hence $M_t^2 - [M]_t = \lim_n (M_t^2 - [M]_t)^{\tau_n}$ is a continuous local martingale. Furthermore, the quadratic variation $[M]_t$ is the unique increasing process with $[M]_0 = 0$ such that $M_t^2 - [M]_t$ is a continuous local martingale. (However, $M_t^2 - [M]_t$ need not be a martingale if M is a local martingale: That require s M to be a square–integrable martingale.)

- We may therefore define $H \bullet M$ as a limit of the integrals $H \bullet M^{\tau_n}$. This means that $H \bullet M$ is a limit of martingales, and hence itself a local martingale. [One needs the fact that $(H \bullet M)^{\tau_n} = (H \bullet M^{\tau_n})$, i.e. that $\int_0^{\tau_n \wedge t} H_s \, dM_s = \int_0^t H_s \, dM_s^{\tau_n}$, which seems obvious.]

- We can also extend our set of integrands: Instead of requiring that H is predictable with $\mathbb{E}[\int_0^\infty H_s^2 \, d[M]_s] < \infty$, it suffices that H be predictable with $\int_0^t H_s^2 \, d[M]_s < \infty$ a.s. for all $t \geq 0$. But then the Itô isometry need not hold, as $||H||_{L^2(M)}$ may be infinite.

- This extended integral thus defined will inherit any of the properties of the L^2–integral which are stable under stopping and taking limits. In particular, it will be linear.

For the remainder of this course, we will pay scant attention to local martingales, and pretend that a local martingale is a martingale, hoping that any problems can be ironed out by localization.

4.2.2 Continuous Semimartingales

We know how to integrate w.r.t. a process A which is locally of finite variation: $(H \bullet A)_t = \int_0^t H_s \, dA_s$ is simply a Riemann–Stieltjes integral (for each $\omega \in \Omega$). We are therefore now able to integrate w.r.t. continuous local martingales and w.r.t. processes of finite variation.

Definition 4.2.1 An adapted càdlàg process X_t is said to be a continuous semimartingale if and only if it has a decomposition

$$X_t = X_0 + M_t + A_t$$

where M is a continuous local martingale with $M_0 = 0$, and A is an adapted finite variation process with $A_0 = 0$.

Observe that every continuous local martingale and every finite variation process is a semimartingale. Moreover, the class of semimartingales is a vector space.

Now integration w.r.t. a semimartingale is easy: If H is predictable, and $X = X_0 + M + A$ is a semimartingale, then we define

$$\int_0^t H_s \, dX_s = \int_0^t H_s \, dM_s + \int_0^t H_s \, dA_s$$

Proposition 4.2.2 *If $X_t = X_0 + M_t + A_t$ is a continuous semimartingale, and H is predictable, then $H \bullet X$ is a semimartingale.*

Proof: We have $H \bullet X = H \bullet M + H \bullet A$, by definition. We know that if M is a local martingale, then $H \bullet M$ is a local martingale. It therefore suffices to prove that $H \bullet A$ is of finite variation, and for that it suffices to show that $H \bullet A$ is a ifference of two increasing processes.

Now clearly if H is non–negative and A is increasing, then $H \bullet A$ is increasing also. Now since A is of finite variation, it is the difference of two increasing processes $A = A_t^{(1)} - A_t^{(2)}$. Furthermore, $H_t = H_t^+ - H_t^-$, where $H_t^+ = \max\{H_t, 0\}$ and $H_t^- = \max\{-H_t, 0\}$. It follows that

$$H \bullet A = (H^+ - H^-) \bullet (A^{(1)} - A^{(2)}) = (H^+ \bullet A^{(1)} + H^- \bullet A^{(2)}) - (H^+ \bullet A^{(2)} + H^- \bullet A^{(1)})$$

is a difference of two increasing processes, as H^+, H^- are non–negative, and $A^{(1)}, A^{(2)}$ are increasing. Hence $H \bullet A$ is of finite variation, and thus

$$H \bullet X = H \bullet M + H \bullet A$$

is a semimartingale decomposition of $H \bullet M$.

$\dashv$

If X is a semimartingale, then the quadratic variation $[X]_t$ can be defined as a limit (in probability) of $\sum_{k=1}^{n(P)}(X_{t_k} - X_{t_{k-1}})^2 = \sum_k (\Delta_k X)^2$ as $\sigma(P) \to 0$.

Similarly, the quadratic covariation of two continuous semimartingales can be defined by

$$[X, Y]_t = \lim_{\sigma(P)\to 0} \sum_{k=1}^{n(P)} (X_{t_k} - X_{t_{k-1}})(X_{t_k} - X_{t_{k-1}})(Y_{t_k} - Y_{t_{k-1}}) = \sum_k \Delta_k X \cdot \Delta_k Y$$

Now recall that a continuous finite variation process has zero quadratic variation. We can say more: If X is a continuous semimartingale, and if A is continuous of finite variation, then $[X, A]_t = 0$ for all t. Indeed

$$\Big|[X, A]_t\Big| \approx \Big|\sum_k \Delta_k \cdot X \Delta_k A\Big| \leq \max_k\{|\Delta_k X|\} \cdot \sum_k |\Delta_k A|$$

Now as $\sigma(P) \to 0$, we have $\max_k\{|\Delta_k X|\} \to 0$, as X is continuous, whereas $\sum_k |\Delta_k A|$ converges to the variation $V_A[0, t]$ of A over $[0, t]$, which is finite. Thus $[X, A]_t = 0 \cdot \mathrm{V}_A[0, t] = 0$.

In particular, if X, Y are continuous semimartingales with decompositions $X_t = X_0 + M_t + A_t$ and $Y_t = Y_0 + N_t + B_t$, then

$$[X, Y] = [M + A, N + B] = [M, N] + [M, B] + [N, A] + [A, B] = [M, N]$$

i.e. the quadratic covariation of two semimartingales equals the quadratic covariation of their local martingale parts.

Note: It is not true that $X_t^2 - [X]_t$ is a martingale when X is a semimartingale. That holds when X is a continuous square–integrable martingale.

It can be proved that semimartingales are the "most general integrators" — This is the Bichteler–Dellacherie Theorem, and you can go and find out what "most general" means yourself. Interestingly, it can also be proved that if a market model has no simple predictable arbitrage opportunities, then the asset prices must be semimartingales — This is a result of Delbaen and Schachermayer.

One common class of semimartingales in mathematical finance is the class of *Itô processes:* A semimartingale X is an Itô process if it has decomposition

$$X_t = X_0 + \int_0^t H_s \, dW_s + \int_0^t K_s \, ds$$

for predictable H and adapted K.

4.2.3 Approximation

Let X, Y be two continuous semimartingales, with decompositions $X_t = X_0 + M_t + A_t$ and $Y_t = Y_0 + N_t + B_t$, and let H, K be predictable processes: As in the L^2-case, we have two approximations:

1. $[X,Y]_t \approx \sum_k \Delta_k X \cdot \delta_k Y$. Since $[X,Y] = [M,N]$, this yields
$$\Delta_k[X,Y] \approx \Delta_k X \cdot \Delta_k Y = \Delta_k M \cdot \Delta_k N$$
and thus
$$\Delta_k[X] \approx (\Delta_k X)^2$$

2. $\int_0^t H_s \, dX_s = \int_0^t H_s \, dM_s + \int_0^t H_s \, dA_s \approx \sum_k H_{t_{k-1}} \Delta_k M + \sum_k H_{t_{k-1}} \Delta_k A = \sum_k H_{t_{k-1}} \Delta_k X$. Thus
$$\Delta_k(H \bullet X) \approx H_{t_{k-1}} \Delta_k X$$

As these were the only things used, it follows exactly as for the L^2–case that

- **(Quadratic Covariation)** $[H\bullet X, K\bullet Y]_t = \int_0^t H_s K_s \, d[X,Y]_t$, i.e. that $[H\bullet X, K\bullet Y] = HK \bullet [X,Y]$.
- **(Associative Law)** $K \bullet (H \bullet X) = (KH) \bullet X$.

4.3 The Itô Formula

Theorem 4.3.1 (One–dimensional Itô Formula) *Suppose that X_t is a continuous semimartingale and that $f \in C^2(\mathbb{R})$ is a twice–continuously differentiable function from $\mathbb{R}$ to $\mathbb{R}$. Then $f(X_t)$ is a continuous semimartingale, with decomposition*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \, dX_s + \frac{1}{2}\int_0^t f''(X_s) \, d[X]_s$$

Proof: Here's another heuristic proof involving approximations that can be justified by some complicated analysis. Recall Taylor's Theorem: If g is C^2, then $g(t+h) = g(t) + g'(t)h + \frac{1}{2}g''(t)h^2 + o(h^2)$.

As ususal, conside a partition $P \equiv 0 = t_0 < t_1 < \cdots < t_{n(P)} = t$ of $[0,t]$. Using the Taylor approximation, we have

$$\begin{aligned} f(X_{t_k}) &\approx f(X_{t_{k-1}}) + f'(X_{t_{k-1}}) \cdot (X_{t_k} - X_{t_{k-1}}) + \frac{1}{2} f''(X_{t_{k-1}}) \cdot (X_{t_k} = X_{t_{k-1}})^2 \\ &= f(X_{t_{k-1}}) + f'(X_{t_{k-1}}) \Delta_k X + \frac{1}{2} f''(X_{t_{k-1}})(\Delta_k X)^2 \end{aligned}$$

Now $(\Delta_k X)^2 \approx \Delta_k[X]$, so

$$f(X_t) - f(X_0) = \sum_k (f(X_{t_k}) - f(X_{t_{k-1}})) \approx \sum_k f'(X_{t_{k-1}})\Delta_k X + \frac{1}{2}\sum_k f''(X_{t_{k-1}})\Delta_k[X]$$

Now as $\sigma(P) \to 0$, the first sum converges to $\int_0^t f'(x_s) \, dX_s$ and the second sum converges to $\int_0^t f''(X_s) \, d[X]_s$.

⊣

Now if $f : \mathbb{R}^m \to \mathbb{R}$ is twice–continuously differentiable, then

$$f(x_1+\Delta x_1, \ldots, x_m+\Delta x_m)) \approx f(x_1, \ldots, x_m)+\sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(x_1, \ldots, x_m)\Delta x_i+\frac{1}{2}\sum_{1\le i,j\le m} \frac{\partial^2 f}{\partial x_i \partial x_j}\Delta x_i \Delta x_j$$

Approximating as above, we get the following theorem.

Theorem 4.3.2 *Suppose that $X = (X^1, \ldots, X^m)$ is an m–tuple of continuous semimartingales, and that $f : \mathbb{R}^m \to \mathbb{R}$ has continuous partial derivatives of second order. Then $f(X^1_t, \ldots X^m_t)$ is a semimartingale, with decomposition*

$$f(X_t) = f(X_0) + \sum_{i=1}^{m} \int_0^t \frac{\partial f}{\partial x_i}(X_t)\ dX^i_t + \frac{1}{2} \sum_{1\le i,j\le m} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)\ d[X^i, X^j]_t$$

Example 4.3.3 Let $X = (B_t, t)$, where B is a Brownian motion, and consider $f(x, t) = x^2 - t$. Then by the Itô formula — and using the fact that t is of finite variation, so that $[B, t] = 0 = 0 = [t, t]$, we see that

$$B^2_t - t = f(B_t, t) = f(B_0, 0) + \int_0^t \frac{\partial f}{\partial x}(B_s, s)\ dB_s + \int_0^t \frac{\partial f}{\partial t}(B_s, s)\ ds + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(B_s, s)\ d[B, B]_s$$

Now $[B, B]_t = [B]_t = t$, so we get

$$B^2_t - t = \int_0^t 2B_s\ dB_s - \int_0^t 1\ ds + \frac{1}{2}\int_0^t 2\ ds = 2\int_0^t B_s\ dB_s$$

And hence

$$\int_0^t B_t\ dB_t = \tfrac{1}{2}(B^2_t - t)$$

as we have already seen.

□

4.4 Differential Notation

Now everything becomes simple...

- We introduce two abbreviations.

  (i) Let

  $$dY_t = H_t\ dX_t \quad \text{be shorthand for} \quad Y_t = Y_0 + \int_0^t H_s\ dX_s$$

  i.e. $d(H \bullet X)_t = H_t\ dX_t$.

  (ii) Further, let

  $$dX_t\ dY_t \quad \text{be shorthand for} \quad d[X, Y]_t$$

Note that $dX_t\ dA_t = 0$ if A is of bounded variation. Thus for any semimartingales X, Y, Z we have $dX_t\ (dY_t\ dZ_t) = 0$ (because $[Y, Z]_t$ is of bounded variation).

- The **distributive laws** become

$$(H_t + K_t)\ dX_t = H_t\ dX_t + K_t\ dX_t \qquad H_t\ d(X+Y)_t = H_t\ dX_t + H_t\ dY_t$$

This looks *obvious*.

- The **associative law** states that if $Y = K \bullet X$, then $H \bullet Y = HK \bullet X$. In differential notation $H_t\ dY_t = H_t K_t\ dX_t$, i.e.

$$H_t(K_t\ dX_t) = (H_t K_t)\ dX_t$$

This looks *obvious*.

- The **covariation** of stochastic integrals: $d[H \bullet X, K \bullet Y]_t$ is (by abbreviation (ii)) $d(H \bullet X)_t\ d(K \bullet Y)_t$. But By abbreviation (i), this can be written as $(H_t\ dX_t)(K_t\ dY_t)$. This looks like it ought to be equal to $H_t K_t\ dX_t\ dY_t$.
However: In abbreviation (ii), $dX_t\ dY_t$ is defined as a *single* object. In the expression $H_t K_t\ dX_t\ dY_t$, the dX_t and dY_t were obtained *separately* from abbreviation (i).
But the covariation of stochastic integrals is given by

$$[H \bullet X, K \bullet Y]_t = \int_0^t H_s K_s\ d[X, Y]_s$$

which abbreviates as $d[H \bullet X, K \bullet Y]_t = H_t K_t d[X, Y]_t = H_t K_t\ dX_t\ dY_t$.
Thus the equation for the covariation of stochastic integrals becomes

$$(H_t\ dX_t)(K_t\ dY_t) = H_t K_t\ dX_t\ dY_t$$

Again: This looks *obvious*.
(But it isn't: The dX, dY on the lefthand side are obtained from abbreviation (i), whereas those on the righthand side are obtained from abbreviation (ii). Thus our notation is *consistent*.)

- The **Itô formula** can now be written as

$$df(\mathbf{X}_t) = \sum_{k \le n} \frac{\partial f}{\partial x_k}(\mathbf{X}_t)\ dX_t^k + \frac{1}{2} \sum_{k,j \le n} \frac{\partial^2 f}{\partial x_k x_j}(\mathbf{X}_t)\ dX_t^k\ dX_t^j$$

This looks like the familiar second order Taylor expansion.

Chapter 5

Girsanov's Theorem and the Martingale Representation Theorem

5.1 Changes of Measure and Girsanov's Theorem

Girsanov's Theorem plays an important rôle in the construction of equivalent martingale measures for processes, and is therefore a cornerstone of the theory of martingale pricing. This section gives a reasonably thorough introduction to it.

5.1.1 Characteristic Functions and Stochastic Exponentials

Suppose that $(X_1, \dots, X_n)$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that the characteristic function $\varphi_{X_1,\dots,X_n} : \mathbb{R}^n \to \mathbb{C}$ is defined by

$$\varphi_{X_1,\dots,X_n}(t_1, \dots, t_n) = \mathbb{E}[e^{i(t_1 X_1 + \cdots + t_n X_n)}] = \mathbb{E}[e^{i\mathbf{t}\cdot\mathbf{X}}]$$

The most important fact about characteristic functions is the following:

Theorem 5.1.1 (Lévy) *Two random vectors have the same distribution if and only if they have the same characteristic function.*

□

The proof of the above theorem may be found in almost any advanced text on probability theory.

As a simple corollary, we have the following useful result:

Corollary 5.1.2 *Two random variables X, Y are independent if and only if $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$.*

Proof: If X, Y are independent, then

$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{i(sX+tY)} = \mathbb{E}[e^{isX} e^{itY}] = \mathbb{E}[e^{isX}] \cdot \mathbb{E}[e^{itY}] = \varphi_X(s)\varphi_Y(t)$$

Conversely, if suppose that $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$. Let $\bar{X}, \bar{Y}$ be independent random variables such that $\bar{X}$ has the same distribution as X, and $\bar{Y}$ as Y. Then $\varphi_{\bar{X}} = \varphi_X$ and $\varphi_{\bar{Y}} = \varphi_Y$. Thus

$$\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t) = \varphi_{\bar{X}}(s)\varphi_{\bar{Y}}(t) = \varphi_{\bar{X},\bar{Y}}(s,t)$$

as $\bar{X}, \bar{Y}$ are independent. It follows that (X, Y) and $(\bar{X}, \bar{Y})$ have the same characteristic function, and thus the same distribution. In particular, X is independent of Y, because $\bar{X}$ is independent of $\bar{Y}$.

$\square$

Theorem 5.1.3 (Kač) *A random vector* $\mathbf{X}$ *is independent of a* σ*–algebra* $\mathcal{G}$ *if and only if*

$$\mathbb{E}[e^{i\mathbf{t}\cdot\mathbf{X}}|\mathcal{G}] = \mathbb{E}[e^{i\mathbf{t}\cdot\mathbf{X}}] \qquad \textit{all } \mathbf{t}$$

Proof: $(\Rightarrow)$ is a basic property of conditional expectation.
$(\Leftarrow)$ We will prove it for the one–dimensional case. Let Y be any $\mathcal{G}$–measurable random variable. Then, using the properties of conditional expectation,

$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{i(sX+tY)}] = \mathbb{E}[\ \mathbb{E}[e^{i(sX+tY)}|\mathcal{G}]\ ] = \mathbb{E}[e^{itY}\mathbb{E}[e^{isX}|\mathcal{G}]\ ] = \mathbb{E}[e^{itY}]\mathbb{E}[e^{isX}] = \varphi_X(s)\varphi_Y(t)$$

Hence X is independent of every $\mathcal{G}$–measurable random variable, and thus independent of $\mathcal{G}$.

$\square$

The *Doléans exponential* of a stochastic process M is defined by:

$$\mathcal{E}(M)_t := e^{M_t - \frac{1}{2}[M]_t}$$

Observe that $\mathcal{E}(M)_t = f(M_t, [M]_t)$ is a function of two processes, namely M_t, and its quadratic variation $[M]_t$, where

$$f(x,y) := e^{x-\frac{1}{2}y}$$

Applying Itô's formula, we see that

$$\begin{aligned} d\mathcal{E}(M)_t &= \mathcal{E}(M)_t\, dM_t - \tfrac{1}{2}\mathcal{E}(M)_t\, d[M]_t + \tfrac{1}{2}\mathcal{E}(M)_t\, d[M]_t \\ &= \mathcal{E}(M)_t\, dM_t \end{aligned}$$

and hence

$$d\mathcal{E}(M)_t = \mathcal{E}(M)_t\, dM_t$$

In particular, $\mathcal{E}(M)_t$ is a local martingale whenever M_t is a local martingale.

A nice application of stochastic exponentials and characteristic functions is given by the following theorem:

Theorem 5.1.4 *(Lévy's Characterization of Brownian Motion) If* $M = (M_t^1, \dots, M_t^d)_t$ *is a continuous* d*–dimensional local martingale such that* $M_0 = 0$ *and* $[M^i, M^j]_t = \delta_{ij}t$*, then* M *is a standard* d*–dimensional Brownian motion.*

Proof: We give the proof for the case $d = 1$, and leave the extension to higher dimensions as an exercise.

Fix $u \in \mathbb{R}$, and define $Y_t := \mathcal{E}(iuM_t) := e^{iuM_t + \frac{1}{2}u^2 t}$. Then $dY_t = iuY_t\, dM_t$, and hence Y_t is a local martingale. It follows easily from the fact that $|Y_t| = e^{\frac{1}{2}u^2 t} < \infty$ that Y is a genuine martingale. Thus $Y_s = \mathbb{E}[Y_t|\mathcal{F}_s]$, i.e.

$$\mathbb{E}[e^{iu(M_t - M_s)}|\mathcal{F}_s] = e^{-\frac{1}{2}u^2(t-s)}$$

The righthandside is deterministic, so taking expectations on both sides, we obtain

$$\mathbb{E}[e^{iu(M_t - M_s)}] = e^{-\frac{1}{2}u^2(t-s)} \qquad \mathbb{E}[e^{iu(M_t - M_s)}|\mathcal{F}_s] = \mathbb{E}[e^{iu(M_t - M_s)}]$$

From the first equation, we see that $(M_t - M_s)$ has the same characteristic function as an $N(0, t-s)$–variables, so that it *is* an $N(0, t-s)$–variable. The second equation, combined with Kač's Theorem, shows that $M_t - M_s$ is independent of $\mathcal{F}_s$.

$\dashv$

Another useful result is the following:

Theorem 5.1.5 *If f is a deterministic function, and W a standard Brownian motion, then*

$$X_t := \int_0^t f(u)\, dW_u$$

is a Gaussian process with independent increments, such that

$$X_t - X_s \sim N\left(0, \int_s^t |f(u)|^2\, du\right)$$

Proof: Define Y_t to be the martingale $Y_t = \mathcal{E}(iuX_t)$, and deduce that

$$\mathbb{E}[e^{iu\int_s^t f(u)\, dW_u}|\mathcal{F}_s] = e^{-\frac{1}{2}\int_s^t |f(u)|^2\, du}$$

noting that $[X]_t = \int_0^t |f(u)|^2\, du$, and using the fact that f is deterministic. Now proceed as in the proof of Lévy's Characterization.

$\dashv$

A question that is important in mathematical finance is the following:

Given that M is a continuous local martingale, when is $\mathcal{E}(M)$ a genuine martingale?

But, a matter of policy, we will gloss over the technical differences between martingales and local martingales. Here, therefore, we will simply state two criteria that partially answer this question. See the book *Stochastic Integration and Differential Equations* by Protter for proofs.

Theorem 5.1.6 (Kazamaki's criterion) *Suppose that M is a continuous local martingale with the property that*

$$\sup_T \mathbb{E}[e^{\frac{1}{2}M_T}] < \infty \qquad \text{where the } \sup \text{ is over all bounded stopping times}$$

Then $\mathcal{E}(M)$ is a UI martingale.

□

Theorem 5.1.7 (Novikov's criterion) *Let M be a continuous local martingale, and assume that*

$$\mathbb{E}[e^{\frac{1}{2}[M]_\infty}] < \infty$$

Then $\mathcal{E}(M)$ is a UI martingale.

□

5.1.2 Changes of Measure

When pricing contingent claims, we use risk-neutral valuation: The $t = 0$ price of a claim X is the *risk–neutral* expectation of its discounted payoff.

$$X_0 = \mathbb{E}_{\mathbb{Q}}[\bar{X}]$$

The measure $\mathbb{Q}$ is not the same as the "real–world" measure $\mathbb{P}$ — we have to change the probability measure.

Suppose that we start with real–world asset dynamics, e.g. a GBM

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

on a filtered space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$, where W_t is a $(\mathbb{F}, \mathbb{P})$–BM. There are two questions that concern us:

- What happens to the dynamics of S_t when we change measures?
- How do we actually go about changing measures?

This is a good time to recall *Bayes' Theorem* for calculating conditional expectations when we change the measure:

Theorem 5.1.8 (Bayes' Theorem)
Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a filtration $\mathcal{F}_n$, and that $\mathbb{Q} << \mathbb{P}$. Let $\xi = d\mathbb{Q}/d\mathbb{P}$ and likelihood process $\xi_t = \mathbb{E}_{\mathbb{P}}[\xi|\mathcal{F}_t]$. Then

(a) For any random variable Z (integrable w.r.t. $\mathbb{P}$ and $\mathbb{Q}$) we have

$$\xi_t \mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[Z\xi|\mathcal{F}_t]$$

(b) If $\mathbb{Q} \approx \mathbb{P}$, then a stochastic process X_t is a martingale under $\mathbb{Q}$ if and only if $\xi_t X_t$ is a martingale under $\mathbb{P}$.

The proof is an exercise:

Exercise 5.1.9 Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a filtration $\mathcal{F}_n$, and that $\mathbb{Q} << \mathbb{P}$. Let $\xi = d\mathbb{Q}/d\mathbb{P}$ and define $\xi_t = \mathbb{E}_{\mathbb{P}}[\xi|\mathcal{F}_t]$, where $\mathbb{E}_{\mathbb{P}}$ refers to *expectation w.r.t. the measure $\mathbb{P}$.*

(a) Show that for any random variable Z (integrable w.r.t. $\mathbb{P}$ and $\mathbb{Q}$) we have

$$\xi_t \mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[Z\xi|\mathcal{F}_t]$$

(b) Show that if $\mathbb{Q} \approx \mathbb{P}$, then a stochastic process X_t is a martingale under $\mathbb{Q}$ if and only if $\xi_t X_t$ is a martingale under $\mathbb{P}$.

[Hint: (a) I'll give you the proof. You justify every step: Let $A \in \mathcal{F}_n$. Then

$$\begin{aligned}\int_A \xi_t \mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t]\, d\mathbb{P} &= \int_A \mathbb{E}_{\mathbb{P}}[\xi \mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t]|\mathcal{F}_t]\, d\mathbb{P} \\ &= \int_A \xi \mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t]\, d\mathbb{P} \\ &= \int_A Z\, d\mathbb{Q} \\ &= \int_A Z\xi\, d\mathbb{P}\end{aligned}$$

(b) Use (a).]

□

Theorem 5.1.10 Girsanov's Theorem for Brownian Motion)
Suppose an d–dimensional process Y has $\mathbb{P}$–dynamics

$$dY_t = \mu_t\, dt + \sigma_t\, dW_t \qquad (t \leq T)$$

where W is a standard d–dimensional $\mathbb{P}$–Brownian motion, $\mu_t(\omega) \in \mathbb{R}^n, \sigma_t(\omega) \in \mathbb{R}^{n\times d}$. Let $\lambda_t(\omega) \in \mathbb{R}^d$ be predictable. Define a measure $\mathbb{Q}$ on $\mathcal{F}_T$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(\lambda \bullet W)_T$$

Assume that Novikov's condition holds:

$$\mathbb{E}\left[e^{\frac{1}{2}\int_0^T ||\lambda_s||^2\, ds}\right] < \infty$$

Then:

(i) $\mathbb{Q}$ *is a probability measure on* $\mathcal{F}_T$.

(ii) $\tilde{W}_t = W_t - \int_0^t \lambda_s\, ds$ *is a $\mathbb{Q}$–Brownian motion.*

(iii) The $\mathbb{Q}$–dynamics of Y are given by

$$dY_t = (\mu_t + \sigma_t\lambda_t)\, dt + \sigma_t\, d\tilde{W}_t$$

Proof: We prove the result for the case $d = 1$, and leave the extension to higher dimensions as an exercise.

By Lévy's characterization, it suffices to show that $\hat{W}_t$ is a continuous local martingale with $[W]_t = t$ under $\mathbb{Q}$. Now certainly

$$d[\hat{W}]_t \text{ "="} (d\hat{W}_t)^2 = d(W_t - \lambda_t\, dt)^2 = (dW_t)^2 = d[W]_t = dt$$

so that $[\hat{W}]_t = [W]_t = t$.

Now let $\xi_t := e^{\int_0^t \lambda_s\, dW_s - \frac{1}{2}\int_0^t ||\lambda_s||^2\, ds} = \mathbb{E}[\xi_T|\mathcal{F}_t]$. Then $\xi_T = \frac{d\mathbb{Q}}{d\mathbb{P}}$ and $(\xi_t)_t$ is a $\mathbb{P}$–martingale, with $d\xi_t = \lambda_t \xi_t\, dW_t$. To prove that $\hat{W}_t$ is a $\mathbb{Q}$–local martingale, it suffices to prove that $\xi_t \hat{W}_t$ is a $\mathbb{P}$–local martingale, by Bayes' Theorem. But

$$\begin{aligned}(\dot{\xi_t \hat{W}_t}) &= \hat{W}_t\, d\xi_t + \xi_t\, d\hat{W}_t + d[\hat{W}_t, \xi_t] \\ &= \hat{W}_t\, d\xi_t + \xi_t\, dW_t - \xi_t\lambda_t\, dt + (dW_t)(\xi_t\lambda_t\, dW_t) \\ &= \hat{W}_t\, d\xi_t + \xi_t\, dW_t\end{aligned}$$

Since both ξ_t and W_t are $\mathbb{P}$–local martingales, it follows that $\xi_t \hat{W}_t$ is the sum of two stochastic integrals w.r.t. $\mathbb{P}$–local martingales, and thus a $\mathbb{P}$–local martingale.

⊣

Remarks 5.1.11 This has important consequences: Suppose, under the "real world" $\mathbb{P}$, we have an asset S the follows a geometric Brownian motion with drift parameter μ and volatility parameter σ. Suppose further that the continuously compounded rate r is constant, and let $\bar{S}_t = e^{-rt} S_t$ denote the discounted asset price. Thus:

$$dS_t = S_t[\mu\, dt + \sigma\, dW_t] \qquad \text{i.e.} \qquad d\bar{S}_t = \bar{S}_t[(\mu - r)\, dt + \sigma\, dW_t]$$

Suppose we now construct a new measure $\mathbb{Q}$ as above. This *Girsanov transformation with kernel* λ adds $\sigma\lambda$ to the drift of $\bar{S}$, but does not change the volatility:

$$d\bar{S}_t = \bar{S}_t[(\mu - r + \sigma\lambda)\, dt + \sigma\, d\tilde{W}_t]$$

We will see that arbitrage–free pricing must be done by computing expectations under a *risk–neutral measure*, or *equivalent martingale measure*. This is a measure under which the discounted asset price process $\bar{S}_t$ is a martingale, and thus has zero drift. Now $\mathbb{Q}$ to be a risk–neutral measure, we must have $(\mu - r + \sigma\lambda) = 0$, i.e.

$$\sigma\lambda = r - \mu \qquad \text{so that} \qquad d\bar{S}_t = \bar{S}_t \sigma\, d\tilde{W}_t$$

This translates to $\lambda = -\frac{\mu - r}{\sigma}$, i.e. the Girsanov kernel is minus the market price of risk.

The fact that a Girsanov transformation does not affect the volatility is also important: It implies that we can use real–world observations to estimate risk–neutral world volatility.

□

5.2 The Martingale Representation Theorem

5.2.1 Motivation

Arbitrage methods only yield a unique price for a contingent claim X when it is possible to replicate X, i.e. when there is a portfolio θ and an initial amount X_0 such that

$$X = X_0 + G_T(\theta) \qquad \text{or equivalently} \qquad \bar{X} = X_0 + \bar{G}_T(\theta)$$

, where $\bar{X}$ refers to the discounted payoff of X. For simplicity, assume that our model has only one risky asset (stock) S. Now in discrete time, the gain on a self–financing previsible portfolio θ involving stock and bank account is given by

$$\bar{G}_T(\theta) = \sum_{k=1}^{T} \theta_k(\bar{S}_k - \bar{S}_{k-1}) = \sum_{k=1}^{T} \theta_k \ \Delta_k \bar{S}$$

where θ_k is the amount of stock held over the interval $[k-1, k]$ and must be $\mathcal{F}_{k-1}$–measurable. Interpolating this expression to continuous time, it is reasonable to model the gain of a continuously traded portfolio θ by a stochastic integral:

$$\bar{G}_T(\theta) = \int_0^T \theta_t \ d\bar{S}_t$$

where θ is predictable, and $\bar{S}$ a semimartingale (so that the stochastic integral is defined.)

Assume now that there is a risk–neutral measure $\mathbb{Q}$, i.e. that the discounted asset price process $\bar{S}_t$ is a $(\mathcal{F}_t, \mathbb{Q})$–martingale. If the stock price dynamics are given by a geometric Brownian motion with volatility parameter σ, then the risk–neutral dynamics are of the form

$$d\bar{S}_t = \sigma \bar{S}_t \ dW_t$$

where W_t is a $(\mathcal{F}_t, \mathbb{Q})$–BM. To hedge an arbitrary contingent claim X, we need to find a predictable process θ such that

$$\bar{X}_T = X_0 + \int_0^T \theta_t \sigma S_t \ dW_t$$

Now define

$$Z_t = \mathbb{E}_{\mathbb{Q}}[\bar{X}_T | \mathcal{F}_t]$$

so that Z_t is a $(\mathcal{F}_t, \mathbb{Q})$–martingale. (Ignore the distinction between martingales and local martingales for the moment.) If we can write the martingale Z as a stochastic integral w.r.t. Brownian motion, i.e. if there exists a predictable process $(H_t)_t$ such that

$$Z_t = Z_0 + \int_0^t H_s \ dW_s$$

then X has a replicating portfolio, namely

$$\theta_t = \frac{H_t}{\sigma S_t}$$

Thus the problem of finding a replicating portfolio for X reduces to finding a way to represent the martingale $Z_t = \mathbb{E}_{\mathbb{Q}}[\bar{X}_T | \mathcal{F}_t]$ as a stochastic integral.

5.2.2 Statement of Main Results

Theorem 5.2.1 (Martingale Representation Theorem)
Let W be a d–dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ be the canonical filtration, augmented to satisfy the usual conditions. Then every

$(\mathbb{F}, \mathbb{P})$–local martingale is representable as a stochastic integral, i.e. for any martingale M there exists a (d–dimensional) predictable process H such that

$$M_t = C + \int_0^t H_s \, dW_s$$

In particular, every $(\mathbb{F}, \mathbb{P})$–martingale has a continuous modification.

This will be a straightforward consequence of the

Theorem 5.2.2 (Itô Representation Theorem)
For any $X \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, there exists a unique predictable process H with $\mathbb{E}[\int_0^T H_t^2 \, dt] < \infty$ such that

$$X = \mathbb{E}[X] + \int_0^\infty H_s \, dW_s$$

The above results, and their proofs below, are taken from *Continuous Martingales and Brownian Motion*, by Revuz and Yor.

5.2.3 Preliminary Technicalities

We aim first to prove the Itô Representation Theorem. We need a few elementary results from Hilbert space theory. Let $(E, \langle \cdot, \cdot \rangle)$ be a Hilbert space.

Definition 5.2.3 A subset $X \subseteq E$ us said to be *total* if the linear span of X (i.e. the set of all linear combinations of vectors in X) is a dense subspace of E.

□

Thus a subset X of E is total if every element of E can be approximated arbitrarily closely by a linear combination of elements of X.

Proposition 5.2.4 *A subset $X \subseteq E$ is total iff whenever*

$$\langle e, x \rangle = 0 \quad \text{for all } x \in X \qquad \Longrightarrow \qquad e = 0$$

Proof: First assume that $X \subseteq E$ is total, so that $E = \mathrm{cl}(\mathrm{span}(X)$, and fix $e \in E$. If $\langle e, x \rangle = 0$ for all $x \in X$, then also $\langle e, f \rangle = 0$ for all $f \in \mathrm{span}(X)$. Choose $x_n \in \mathrm{span}(X)$ such that $x_n \to e$. Then $\langle e, x_n \rangle \to \langle e, e \rangle$, so $\langle e, e \rangle = ||e||^2 = 0$. hence $e = 0$.

Conversely, assume that $X \subseteq E$ satisfies the stated condition. Let $F = \mathrm{cl}(\mathrm{span}(X)$. We must prove that $E = F$. So let $e \in E$ be arbitrary, and write $e = e_{||} + e_\perp$, where $e_{||} \in F, e_\perp \perp F$. Then $\langle e_\perp, x \rangle = 0$ for all $x \in X$, so that $e = e_{||} \in F$. Hence $E \subseteq F$, as required.

⊣

The Itô Representation Theorem is a statement about the Hilbert space $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. We begin by finding a convenient total subset of this space. For simplicity assume that W is a 1–dimensional Brownian motion. Let $\mathcal{S}$ be the set of *deterministic* step functions on $\mathbb{R}^+$, i.e. those functions f which can be written

$$f = \sum_{i=1}^n \lambda_i I_{(t_{i-1}, t_i]}$$

Lemma 5.2.5 *The set $\mathcal{E}(\mathcal{S}) = \{\mathcal{E}(f)_\infty = e^{\int_0^\infty f(t)\,dW_t - \frac{1}{2}\int_0^\infty f(t)\,dt} : f \in \mathcal{S}\}$ of Doléans exponentials is total in $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$.*

Proof: Let $Y \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ be such that $\mathbb{E}[\mathcal{E}(f)_\infty \cdot Y] = 0$ for all $f \in \mathcal{S}$. We want to show that $Y = 0$ a.s. Fix $0 = t_0 \le t_1 \le \cdots \le t_n$, and define $\varphi : \mathbb{C}^n \to \mathbb{C}$ by

$$\varphi(z_1, \dots, z_n) = \mathbb{E}\left[e^{\sum_{j=1}^n z_j (W_{t_j} - W_{t_{j-1}})} \cdot Y\right]$$

Now if $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, and $f = \sum_{j=1}^n \lambda_j I_{(t_{j-1}, t_j]}$, then

$$\varphi(\lambda_1, \dots, \lambda_n) = e^{\frac{1}{2}\int_0^\infty f(t)^2\,dt}\mathbb{E}[\mathcal{E}(f)_\infty \cdot Y] = 0$$

It is now not too hard to believe that $\varphi(i\lambda_1, \dots, i\lambda_n) = 0$ for all $\lambda_1, \dots, \lambda_n$ (To be precise: Noting that W is a Gaussian process, it is not hard to see that φ is analytic. Since $\varphi = 0$ on $\mathbb{R}^n$, we must have $\varphi = 0$ on $\mathbb{C}^n$, by analytic continuation.)

The expression $\varphi(i\lambda_1, \dots, i\lambda_n) = \mathbb{E}\left[e^{i\sum_{j=1}^n \lambda_j (W_{t_j} - W_{t_{j-1}})} \cdot Y\right]$ looks a little like a characteristic function, i.e. a Fourier transform. To make this explicit, define a signed measure μ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ by $\mu(A) = \mathbb{E}\left[I_A(W_{t_1} - W_{t_0}, \dots, W_{t_n} - W_{t_{n-1}}) \cdot Y\right]$, i.e. $\mu = \nu X^{-1}$, where $d\nu = Y\,d\mathbb{P}$ and $X = (W_{t_1} - W_{t_0}, \dots, W_{t_n} - W_{t_{n-1}})$. Then

$$\varphi(i\lambda_1, \dots, i\lambda_n) = \int e^{i\langle \lambda, x\rangle}\,\mu(dx) \qquad \lambda = (\lambda_1, \dots, \lambda_n)$$

is the Fourier transform of the measure μ. Since it is identically zero, by the Fourier inversion theorem (analogous to Lévy inversion), the measure μ is identically zero. Now clearly $\sigma(X) = \sigma(W_{t_1}, \dots, W_{t_n})$, so the measure ν is zero on $\sigma(W_{t_1}, \dots, W_{t_n})$, for all $0 = t_0 \le t_1 \le \cdots \le t_n$.

Now let $\{q_n : n \in \mathbb{N}\}$ enumerate a dense subset of $\mathbb{R}^n$, and let $\mathcal{H}_n = \sigma(W_{q_1}, \dots, W_{q_n})$. We've just proved that ν is zero on each $\mathcal{H}_n$. Fix an arbitrary bounded $\mathcal{F}_\infty$–measurable variable Z, and let $Z_n = \mathbb{E}[Z|\mathcal{H}_n]$. By the martingale convergence theorem, $Z_n \to Z$ a.s. and in L^1. Hence, since Z is bounded, $Z_n Y \to ZY$ a.s. and in L^1. But since ν is zero on $\mathcal{H}_n$, we have $\mathbb{E}[Z_n Y] = \int Z_n\,d\nu = 0$ for all n. It follows that $\mathbb{E}[ZY] = 0$ for all bounded $\mathcal{F}_\infty$–measurable R.V.'s Z. Take $Z = Y \wedge N$ for sufficiently big N to conclude $Y = 0$ a.s.

$\dashv$

Further recall the following: In constructing the stochastic integral, we made use of the following isometry for square–integrable martingales M:

$$||H||_{L^2(M)} = ||H \bullet M||_{\mathcal{M}^2} \qquad \text{i.e.} \qquad \mathbb{E}\left[\int_0^\infty H_t^2\,d[M]_t\right] = \mathbb{E}\left[\left(\int_0^\infty H_t\,dM_t\right)^2\right]$$

5.2.4 Proof of Main Results

Proof of Itô Representation Theorem: Uniqueness is a simple consequence of the Itô isometry: Indeed, if $(H \bullet W)_\infty = (K \bullet W)_\infty$, then $||H - K||_{L^2(W)} = ||(H - K) \bullet W||_{\mathcal{M}^2} = 0$, so that $H = K$ in $L^2(W)$.

Denote by $\mathcal{H}$ the set of all $X \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ which can be represented in the form

$$X = \mathbb{E}[X] + \int_0^\infty H_t\,dW_t$$

for some predictable H with $\mathbb{E}[\int_0^\infty H_t^2\ dt] < \infty$. By linearity of the integral, $\mathcal{H}$ is a subspace of L^2. Moreover, $\mathcal{E}(\mathcal{S}) \subseteq \mathcal{H}$: Indeed, if $f = \sum_{j=1}^n \lambda_j I_{(t_{j-1},t_j]} \in \mathcal{S}$, then $d\mathcal{E}(f)_t = \mathcal{E}(f)_t f(t)\ dW_t$ — the well–known SDE satisfied by Doléans exponentials — so that $\mathcal{E}(f)_\infty = 1 + \int_0^\infty \mathcal{E}(f)_t f(t)\ dW_t$.

Since $\mathcal{E}(S)$ is total and $\mathcal{E}(\mathcal{S}) \subseteq \mathcal{H}$, it suffices to show that $\mathcal{H}$ is closed. So let $(X_n)_n$ be a Cauchy sequence in L^2, i.e.

$$\mathbb{E}[(X_n - X_m)^2] = E[X_n - X_m]^2 + ||(H^n - H^m) \bullet W||^2_{\mathcal{M}^2} \to 0 \quad \text{as} \quad n, m \to \infty$$

We can conclude two things: (i) $\mathbb{E}[X_n - X_m]^2 \to 0$, from which we deduce that $(\mathbb{E}[X_n])_n$ is Cauchy, hence convergent, and (ii) $||(H^n - H^m) \bullet W||^2_{\mathcal{M}^2} \to 0$, from which we deduce that $(H^n)_n$ is a Cauchy sequence in $\mathcal{L}^2(W)$, by the isometry, and hence convergent to some predictable H. It is now easy to see that the sequence $(X_n)_n$ converges to

$$\lim_n \mathbb{E}[X_n] + \int_0^\infty H_t\ dW_t$$

which belongs to $\mathcal{H}$. Since every Cauchy sequence in $\mathcal{H}$ converges, it is closed.

$\dashv$

Proof of Martingale Representation Theorem: First suppose that M is a square–integrable martingale; then there is a unique predictable H such that

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty H_t\ dW_t$$

Then $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t] = M_0 + \int_0^t H_s\ dW_s$. It follows that square–integrable martingales have continuous versions.

Next suppose that M is a uniformly integrable martingale (so that it converges to M_∞ a.s. and in L^1). Since $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is dense in $L^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$, we can choose random variables $M_\infty^n \in L^2$ such that $\mathbb{E}[|M_\infty^n - M_\infty|] \to 0$. Define martingales M^n by $M_t^n = \mathbb{E}[M_\infty^n|\mathcal{F}_t]$. Then the M^n are square–integrable, and thus continuous. By Doob's maximal inequality, we see that

$$\mathbb{P}\left[\sup_t |M_t - M_t^n| > \lambda\right] \leq \lambda^{-1}\mathbb{E}[|M_\infty^n - M_\infty|] \to 0$$

By Borel–Cantelli Lemma we can pick a subsequence $(M^{n_k})_k$ which converges uniformly to M. Hence M has a continuous version.

Finally, let M be a local martingale. Then there exists a sequence of stopping times $T_n \uparrow \infty$ such that each stopped martingale M^{T_n} is uniformly integrable, hence continuous. Modifying the T_n, if necessary, we may assume that each M^{T_n} is bounded, hence square–integrable. By the first part of the proof, we see that

$$M_t^{T_n} = M_0 + \int_0^{T_n \wedge s} H_s^n\ dW_s$$

for some unique predictable H^n. Uniqueness ensure furthermore that if $m \leq n$, then H^n, H^m coincide on $(0, T_m]$, and we may denote this common value by H. The result follows.

$\dashv$

Chapter 6

SDEs and PDEs

6.1 SDEs: Existence and Uniqueness of Solutions

An ODE is some functional relationship

$$f(t, x'(t), x''(t), \dots) = 0$$

involving, say, time t and an unknown function x and its derivatives.

Example 6.1.1 Consider the following population growth model

$$\frac{dN}{dt} = a(t)N(t) \qquad N(0) = N_0$$

where $N(t)$ is the size of the population at time t. Of course, this is easy to solve:

$$N(t) = N_0 e^{\int_0^t a(s)\, ds}$$

Here $a(t)$, the relative growth rate, is deterministic. However, we can easily imagine there to be "noise" in the system, i.e.

$$a(t) = r(t) + \text{“random noise”}$$

If we interpret ("noise")$\cdot$ dt to be some random perturbation $\sigma(t)\, dW_t$, where W_t is a standard Brownian motion, then we obtain

$$dN(t) = r(t)N(t)\, dt + \sigma(t)N(t)\, dW_t$$

We choose to interpret this in the Itô sense:

$$N(t) = N_0 + \int_0^t r(s)N(s)\, ds + \int_0^t \sigma(s)N(s)\, dW_s$$

If r, σ are constants, this becomes our old friend geometric Brownian motion

$$\frac{dN_t}{N_t} = r\, dt + \sigma\, dW_t$$

whose solution we know:

$$N(t) = N(0)e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t}$$

as you can easily verify by Itô's formula.

□

Definition 6.1.2 An Itô diffusion is an SDE of the form

$$dX_t = b(t, X_t)\ dt + \sigma(t, X_t)\ dW_t$$

where

$$b : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$$
$$\sigma : \mathbb{R}^+ \times \mathbb{R}^n \to R^{n\times m}$$

Here, X_t is an n–dimensional stochastic process, and W_t is an m–dimensional Brownian motion. Thus

$$\begin{pmatrix} dX_t^1 \\ \vdots \\ dX_t^n \end{pmatrix} = \begin{pmatrix} b^1(t, X_t^1, \ldots, X_t^n) \\ \vdots \\ b(t, X_t^1, \ldots, X_t^n) \end{pmatrix}\ dt + \begin{pmatrix} \sigma_{11}(t, X_t^1, \ldots, X_t^n) & \ldots & \sigma_{1m}(t, X_t^1, \ldots, X_t^n) \\ \vdots & \vdots & \vdots \\ \sigma_{n1}(t, X_t^1, \ldots, X_t^n) & \ldots & \sigma_{nm}(t, X_t^1, \ldots, X_t^n) \end{pmatrix} \begin{pmatrix} dW_t^1 \\ \vdots \\ dW_t^m \end{pmatrix}$$

which is to be interpreted as a *system of stochastic integral equations*

$$X_t^i = X_0^i + \int_0^t b^i(s, X_s^1, \ldots, X_s^n)\ ds + \sum_{k=1}^m \int_0^t \sigma_{ik}(s, X_s^1, \ldots, X_s^n)\ dW_s^k$$

We denote the SDE above by SDE(σ, b).

□

As with ODE's, there is a theorem which guarantees that, under certain conditions, solutions solutions exist and are unique. Before we state this result, it's important to note that there are two common notions of *solution* to an SDE. Given SDE(σ, b): a *solution* is a triple $(X_t, B_t, \mathcal{F}_t^*)$ such that

(i) B_t is an $\mathcal{F}_t^*$–Brownian motion;

(ii) X_t satisfies

$$X_t = X_0 + \int_0^t b(s, X_s)s\ ds + \int_0^t \sigma(s, X_s)\ dB_s$$

If X_t is adapted to the filtration generated by B_t, then X_t is called a *strong* solution. In essence, given a Brownian motion B_t, we can the construct a solution to SDE(σ, b) from this Brownian motion B_t. However, it may be impossible to solve SDE(σ, b) using a given Brownian motion, but nevertheless possible to solve it by constructing X_t and a different Brownian motion — i.e. it is necessary to construct X_t and B_t simultaneously. In that case, we call the solution a *weak* solution.

We are mainly interested in *strong solutions*. We say that a solution is (pathwise) *unique* if given any two solutions $(X_t, B_t, \mathcal{F}_t)$ and $(X_t', B_t, \mathcal{F}_t')$ to SDE(σ, b) with $X_0 = X_0' = x$, driven by the *same* Brownian motion B_t, we have, with probability one,

$$X_t = X_t' \quad \text{for all } t \geq 0$$

Without proof, we state:

Theorem 6.1.3 (Existence and Uniqueness Theorem)
Let $T > 0$. Given the following:

- *An m–dimensional Brownian motion W_t;*
- *An SDE SDE(σ, b);*
- *An n–dimensional random variable Z independent of $(W_t)_{t \leq T}$ with $\mathbb{E}Z^2 < \infty$ (in particular, Z may be constant).*

Suppose that there is a constant C such that

(i) The following local Lipschitz condition *holds: For all $x, y \in \mathbb{R}^n$ and all $0 \leq t \leq T$ we have*

$$||b(t,x) - b(t,y)|| \leq C||x - y||$$
$$||\sigma(t,x) - \sigma(t,y)|| \leq C||x - y||$$

(ii) The following linear growth condition *holds: For all $x \in \mathbb{R}^n, 0 \leq t \leq T$*

$$||b(t,x)|| + ||\sigma(t,x)|| \leq K(1 + ||x||)$$

Then there exists a unique strong solution *X_t to the SDE*

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$$
$$X_0 = Z$$

Moreover

1. *X_t is adapted to the natural filtration (augmented) generated by Z and W_t.*
2. *X_t has continuous sample paths.*
3. *X_t is a (strong) Markov process.*
4. *$\mathbb{E}\int_0^T ||X_t||^2\, dt < \infty$.*

$\square$

The above is not the best possible theorem. In the 1–dimensional case, particularly, it can be strengthened considerably.

6.2 The Linear SDE

In this section we "solve" the one–dimensional linear SDE

$$dX_t = [b_1(t)X_t + b_2(t)]\, dt + [\sigma_1(t)X_t + \sigma_2(t)]\, dW_t$$

where b_i, σ_i are deterministic continuous functions (and thus bounded on compact intervals). The Existence and Uniqueness Theorem guarantees the existence of a unique strong solution. We will solve this SDE in three steps.

6.2.1 The Linear SDE with additive noise only

Consider

$$dX_t = [b_1 X_t + b_2]\, dt + \sigma_2\, dW_t$$

(i.e. $\sigma_1 = 0$.) An ordinary linear DE $x'(t) = b(t)x(t) + u(t)$ is solved using an integrating factor $y(t) = e^{-\int b(t)\, dt}$ which reduce the problem to $(x(t)y(t))' = u(t)$, which can be solved by integrating both sides. We try the same trick here. So put

$$y(t) = e^{-\int_0^t b_1(s)\, ds} \qquad Y_t = y(t)X_t$$

Then by Itô's formula

$$\begin{aligned} dY_t &= -b_1(t)y(t)X_t\, dt + y(t)\, dX_t \\ &= y(t)[-b_1(t)X_t + b_1(t)X_t + b_2(t)]\, dt + \sigma_2(t)\, dW_t \\ &= y(t)b_2(t)\, dt + y(t)\sigma_2(t)\, dW_t \end{aligned}$$

Note that $Y_0 = X_0$, because $y(0) = 1$. Integrating the above equation, we obtain

The solution to the SDE

$$dX_t = [b_1(t)X_t + b_2(t)]\, dt + \sigma_2(t)\, dW_t$$

is given by

$$X_t = [y(t)]^{-1}\left[X_0 + \int_0^t b_2(s)y(s)\, ds + \int_0^t \sigma_2(s)y(s)\, dW_s\right]$$

where $y(t) = e^{-\int_0^t b_1(s)\, ds}$.

Example 6.2.1 The Langevin Equation

$$dX_t = cX_t\, dt + \sigma\, dW_t$$

Here $b_1(t) = c =$ constant, $b_2(t) = 0$, $\sigma_1(t) = 0$, and $\sigma_2(t) = \sigma =$ constant.
Thus the integrating factor is $y(t) = e^{-\int_0^t c\, ds} = e^{-ct}$, and hence

$$X_t = e^{ct}X_0 + \sigma \int_0^t e^{c(t-s)}\, dW_s$$

Now recall that if $h(t)$ is deterministic, then $Z = \int_0^t h(s)\ dW_s$ is Gaussian, with $Z \sim N(0, \int_0^t h(s)^2\, ds)$ (using the Itô isometry). It follows that we know the distribution of X_t: X_t is normally distributed with mean $e^{ct}\mathbb{E}[X_0]$ and variance $\sigma^2 \int_0^t e^{2c(t-s)}\, ds = \sigma^2 \frac{e^{2ct}-1}{2c}$.

□

Example 6.2.2 The Vasicek Model:
This is a short rate model:

$$dr_t = c[\mu - r_t]\, dt + \sigma\, dW_t$$

where μ, σ are constant. Thus $b_1(t) = -c$, $b_2(t) = c\mu$, $\sigma_1(t) = 0$ and $\sigma_2(t) = \sigma$. The integrating factor is $y(t) = \exp(\int_0^t c\, ds) = e^{ct}$, and so

$$\begin{aligned} r_t &= e^{-ct}\left[r_0 + \int_0^t c\mu e^{cs}\, ds + \int_0^t \sigma e^{cs}\, dW_s\right] \\ &= r_0 e^{-ct} + \mu[1 - e^{-ct}] + \sigma \int_0^t e^{-c(t-s)}\, dW_s \end{aligned}$$

Thus r_t is normally distributed with mean $r_0 e^{-ct} + \mu(1 - e^{-ct})$ and variance $\sigma^2 \frac{1-e^{-2ct}}{2c}$. As $t \to \infty$, the short rate "forgets" its initial value r_0 and lies approaches a normally distributed random variable with mean μ and variance $\frac{\sigma^2}{2c}$.

□

6.2.2 The Homogeneous Linear SDE with Multiplicative Noise

Next, we consider

$$dX_t = b_1(t)X_t\, dt + \sigma_1(t)X_t\, dW_t$$

i.e. $b_2(t) = 0 = \sigma_2(t)$. If b_1, σ_1 are constants, the above SDE is simply geometric Brownian motion. So try the following: Define $Y_t = \ln X_t$. Then

$$\begin{aligned} dY_t &= \frac{1}{X_t}\, dX_t - \frac{1}{2X_t^2}\, d[X]_t \\ &= [b_1(t) - \frac{1}{2}\sigma_1(t)^2]\, dt + \sigma_1(t)\, dW_t \end{aligned}$$

and thus

$$Y_t = Y_0 + \int_0^t b_1(s) - \frac{1}{2}\sigma_1(s)^2\, ds + \int_0^t \sigma_1(s)\, dW_s$$

so that

> The solution to the SDE
>
> $$dX_t = b_1(t)X_t\, dt + \sigma_1(t)X_t\, d_t$$
>
> is
>
> $$\begin{aligned} X_t &= X_0 e^{\int_0^t \sigma_1(s)\, dW_s + \int_0^t b_1(s) - \frac{1}{2}\sigma_1(s)^2\, ds} \\ &= X_0 \mathcal{E}\left(\int b_1\, dt + \int \sigma_1\, dW_t\right) \end{aligned}$$

In the case of additive noise only, we saw that the solutions are Gaussian processes. In the case of homogeneous SDE's with multiplicative noise, solutions are lognormally distributed.

6.2.3 The General Linear SDE

Given

$$dX_t = [b_1(t)X_t + b_2(t)]\, dt + [\sigma_1(t)X_t + \sigma_2(t)]\, dW_t$$

let Y_t be the solution to the corresponding homogeneous SDE

$$dY_t = b_1(t)X_t\, dt + \sigma_1(t)X_t\, dW_t \qquad Y_0 = 1$$

An application of Itô's formula shows that

$$\begin{aligned} d\left(\frac{X}{Y}_t\right) &= \frac{1}{Y_t}\,dX_t - \frac{X_t}{Y_t^2}\,dY_t - \frac{1}{Y_t^2}\,d[X,Y]_t + \frac{2X_t}{2Y_t^3}\,d[Y]_t \\ &= \frac{1}{Y_t}\left[(b_2(t) - \sigma_1(t)\sigma_2(t))\,dt + \sigma_2(t)\,dW_t\right] \end{aligned}$$

Integrate:

$$\frac{X_t}{Y_t} = \frac{X_0}{Y_0} + \int_0^t \frac{b_2(s) - \sigma_1(s)\sigma_2(s)}{Y_s}\,ds + \int_0^t \frac{\sigma_2(s)}{Y_s}\,dW_s$$

and thus

The solution to the SDE

$$dX_t = (b_1(t)X_t - b_2(t))\,dt + (\sigma_1(t)X_t + \sigma_2(t))\,d_t$$

is given by

$$X_t = Y_t\left[X_0 + \int_0^t \frac{b_2(s) - \sigma_1(s)\sigma_2(s)}{Y_s}\,ds + \int_0^t \frac{\sigma_2(s)}{Y_s}\,dW_s\right]$$

where Y_t solves the corresponding homogeneous SDE with $Y_0 = 1$:

$$Y_t = e^{\int_0^t \sigma_1(s)\,dW_s + \int_0^t b_1(s) - \frac{1}{2}\sigma_1(s)^2\,ds}$$

6.3 Solving PDEs Probabilistically

6.3.1 Solving the Heat Equation Probabilistically

We show here how it is possible to solve certain PDE's by running a Brownian motion. For simplicity's sake consider the d–dimensional heat equation, with initial conditions:

$$\begin{aligned} u_t(t,x) &= \frac{1}{2}\Delta u \\ u(0,x) &= f(x) \end{aligned} \tag{*}$$

where $u : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is $C^{1,2}$ (and $x = (x^1, \ldots, x^d)$). For example, in two dimensions,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2}\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 y}{\partial y^2}\right] \\ u(0,x,y) &= f(x,y) \end{aligned}$$

We will also write $\nabla u = (\frac{\partial u}{\partial x^1}, \ldots, \frac{\partial u}{\partial x^d})$.

To begin with, we note that

Proposition 6.3.1 *If u satisfies $(*)$, then $M_s = u(t-s, B_s)$ is a local martingale on $[0,t)$. Here B_t is a d–dimensional Brownian motion (not necessarily null at 0).*

Proof: Let $X_s^0 = t - s$ and $X_s^i = B_s^i$ for $1 \leq i \leq d$. Further define $Y_s = u(X_s^0, \ldots, X_s^d) = u(t-s, B_s)$, and apply Itô's formula to obtain

$$dY_s = -u_t(t-s, B_s)\; ds + \nabla u(t-s, B_s) \cdot\; dB_s + \frac{1}{2}\Delta u(t-s, B_s)\; ds$$

where we use the fact that $d[B^i]_s = ds, d[B^i, B^j]_s = 0$ for $i \neq j$ both ≥ 1. Hence

$$M_s = u(t-s, B_s) = u(t, B_0) + \int_0^s \nabla u(t-r, B_r) \cdot\; dB_r$$

is a local martingale.

$\square$

Now let B_t be a Brownian motion starting at $x \in \mathbb{R}^d$. Then

$$u(t,x) = M_0 = \mathbb{E}^x[M_t] = \mathbb{E}^x[u(0, B_t)] = \mathbb{E}^x[f(B_t)]$$

where $\mathbb{E}^x$ denotes the expectation under a measure where B_t starts at x. Thus we can solve the heat equation as follows:

(i) Start a Brownian motion from x, and let it run for a time t.

(ii) Plug the value B_t into the initial condition f to obtain a random variable $f(B_t)$.

(iii) $u(t,x)$ is the expectation of $f(B_t)$: $u(t,x) = \mathbb{E}^x[f(B_t)]$

Thus it is possible to solve the heat equation by running a Brownian motion. With some modifications, we can solve many parabolic and elliptic problems by running some stochastic process and taking expectations. We now begin the process of making this precise.

6.3.2 The Black-Scholes PDE: A Heuristic Approach

Using Itô's formula, it is not hard to derive a partial differential equation for European–style derivatives.

Consider again market with a share S_t whose price process is given by a geometric Brownian motion, i.e. satisfies the SDE

$$dS = \mu S\; dt + \sigma S\; dB_t$$

Let the risk–free interest rate be r, and let A_t be the riskless bank account, with dynamics

$$dA_t = r A_t\, dt$$

Let $V(t, S_t)$ be European–style derivative whose value depends on both the share price and time. Consider a portfolio Π which contains 1 derivative, and n shares, i.e. its value is

$$\Pi_t = V_t + nS_t$$

A small amount of time dt later, the share price has changed. The value of the portfolio changes by

$$d\Pi_t = dV_t + n\, dS_t$$

By Itô's Formula,

$$\begin{aligned} dV_t &= \frac{\partial V}{\partial t}\,dt + \frac{\partial V}{\partial S}\,dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\,dS^2 \\ &= \left(\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)dt + \sigma S\frac{\partial V}{\partial S}\,dB_t \end{aligned}$$

Hence

$$\begin{aligned} d\Pi_t = &\left(\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + n\mu S\right)dt \\ &+ \sigma S\left(\frac{\partial V}{\partial S} + n\right)dB_t \end{aligned}$$

Thus

$$\begin{aligned} d\Pi_t = &\left(\frac{\partial V}{\partial t} + \mu S\left[\frac{\partial V}{\partial S} + n\right] + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)dt \\ &+ \sigma S\left[\frac{\partial V}{\partial S} + n\right]dB_t \end{aligned}$$

Now if we take $n = -\frac{\partial V}{\partial S}$ (i.e. the portfolio is short $-\frac{\partial V}{\partial S}$ shares), then the portfolio is unaffected by the random changes in stock prices:

$$d\Pi_t = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)dt$$

Thus, for a brief moment, the portfolio is risk–free. By a *no–arbitrage argument*, it must *earn the same return as the risk–free bank account*[1], i.e.

$$d\Pi_t = r\Pi_t\,dt = r\left(V - \frac{\partial V}{\partial S}S\right)dt$$

Equating (6.3.2) and (6.3.2), we get

$$\boxed{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0}$$

This is the famous **Black–Scholes** PDE. It is a second–order parabolic PDE, i.e. essentially a heat equation. Most of the PDE's encountered in finance are of a similar type.

Note that if a portfolio contains $\frac{\partial V}{\partial S}$ shares, then the change in the portfolio value is the same as the change in the value of the derivative. The quantity $\frac{\partial V}{\partial S}$ is called the *delta* of the derivative. One can thus *synthetically replicate* any European style derivative with underlying share S by holding, at any time, delta–many shares. This procedure is called *delta hedging.*

Consider a European call option C on a share S with strike K and maturity T. The volatility of the underlying share S is σ and the risk–free rate is r. To find the value of the call option, we must solve the following boundary value problem:

$$\begin{cases} \dfrac{\partial C}{\partial t} + \dfrac{1}{2}\sigma^2 S^2\dfrac{\partial^2 C}{\partial S^2} + rS\dfrac{\partial C}{\partial S} - rC = 0 \\ C(T) = \max\{S_T - K, 0\} \end{cases}$$

[1] This is the *crux* of the argument!

6.3.3 The Feynman–Kac Theorem

Consider an (n–dimensional) SDE

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$$

with d sources of noise. Thus μ is an n–vector, and σ an $n \times d$–matrix.

Let $f(t, x^1, \ldots, x^n) : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ be a $C^{1,2}$–function. Let $\nabla_x f(t, x^1, \ldots, x^n) = (\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n})$, and let $C = \sigma\sigma^{tr}$. Note that C is an $n \times n$–matrix, and that $C_{ij} = \sigma_i \cdot \sigma_j$, where σ_i is the i^{th} row of the matrix σ.

By Itô's formula

$$df(t, X_t) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \sum_{j=1}^d \sigma_{ij}\, dW_t^j + \left[\frac{\partial f}{\partial t} + \sum_{i=1}^n \mu_i \frac{\partial f}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^n C_{ij} \frac{\partial^2 f}{\partial x^i \partial x^j} \right] dt$$

i.e.

$$df(t, X_t) = \nabla_x f \cdot \sigma\, dW_t + \left[\frac{\partial f}{\partial t} + \mu \cdot \nabla_x f + \frac{1}{2} \sum_{i,j} C_{ij} \frac{\partial^2 f}{\partial x^i \partial x^j} \right] dt$$

Definition 6.3.2 The *infinitesimal generator* $\mathcal{A}$ of a diffusion X_t (satisfying $dX_t = \mu\, dt + \sigma\, dW_t$) is defined by

$$\mathcal{A} f(t, x^1, \ldots, x^n) = \sum_{i=1}^n \mu_i \frac{\partial f}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^n C_{ij} \frac{\partial^2 f}{\partial x^i \partial x^j}$$

where $C = \sigma\sigma^{tr}$.

□

Thus

$$df(t, X_t) = \left[\frac{\partial f}{\partial t} + \mathcal{A} f \right] dt + \nabla_x f \cdot \sigma\, dW_t$$

Consider now the following *Cauchy problem*:

$$\begin{gathered} \frac{\partial V}{\partial t}(t, x) + \mu(t, x) \frac{\partial V}{\partial x} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 V}{\partial x^2} = 0 \\ V(T, x) = \Phi(x) \end{gathered} \tag{*}$$

We will solve this PDE probabilistically, by running a diffusion.

To solve it, we must find $V(t, x)$, for $0 \le t \le T$ and $x \in \mathbb{R}$. So fix $t \le T$ and $x \in \mathbb{R}$, and define a 1–dimensional diffusion X to be the solution of

$$\begin{aligned} dX_s &= \mu(s, X_s)\, ds + \sigma(s, X_s)\, dW_s \qquad t \le s \le T \\ X_t &= x \end{aligned}$$

Thus X starts running at time t from point x. The infinitesimal generator of X is

$$\mathcal{A} = \mu(t, x) \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2}{\partial x^2}$$

Thus $(*)$ is just

$$\frac{\partial V}{\partial t} + \mathcal{A}V = 0$$
$$V(T,x) = \Phi(x)$$

Applying Itô's formula to the process $V(t, X_t)$ yields

$$dV_t = \frac{\partial V}{\partial x}\sigma\, dW_t + \left[\frac{\partial V}{\partial t} + \mathcal{A}V\right]\, dt$$

Since the term in brackets is zero (Cauchy problem), we see that

$$V(T, X_T) = V(t, X_t) + \int_t^T \sigma(s, X_s)\frac{\partial V}{\partial x}(s, X_s)\, dW_s$$

and thus, noting that $V(T, X_T) = \Phi(X_T)$, that $X_t = x$, and taking expectations on both sides, that

$$V(t,x) = \mathbb{E}^{t,x}[\Phi(X_T)]$$

Thus the solution $V(t,x)$ to the Cauchy problem can be obtained by running an SDE from point x at time t, waiting until time T, and finding the average value of the random variable $\Phi(X_T)$. The superscripts t, x on $\mathbb{E}^{t,x}$ simply denote that $X_t = x$.

Now $(*)$ is not quite the Black–Scholes equation, which has an additional term. However, this can be removed. We obtain the following general, multidimensional, vesrion:

Theorem 6.3.3 (Feynman–Kac Theorem)
Given

- *A (column) vector–valued function* $\mu : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$*;*
- *A matrix–valued function* $\sigma : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$*;*
- *A matrix–valued function* C *which is of the form* $C = \sigma\sigma^{tr}$*;*
- *A real–valued function* $r : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$*;*
- *A real–valued function* $\Phi : \mathbb{R}^n \to \mathbb{R}$*;*

Given a solution $V(t,x)$ *to the boundary value problem*

$$\frac{\partial V}{\partial t} + \sum_{i=1}^n \mu_i \frac{\partial V}{\partial x^i} + \frac{1}{2}\sum_{i,k=1}^n C_{ij}\frac{\partial^2 V}{\partial x^i \partial x^j} + rV = 0$$
$$V(T,x) = \Phi(x)$$

and assuming sufficient integrability, we can calculate $V(t,x)$ *as follows:*

(i) Fix $t \le T$ *and* $x \in \mathbb{R}$*;*

(ii) Let X_s *be the solution to*

$$dX_s = \mu(s, X_s)\, ds + \sigma(t, X_s)\, dW_s \qquad t \le s \le T$$
$$X_t = x$$

Then

$$V(t,x) = \mathbb{E}^{t,x}\left[e^{\int_t^T r(s,X_s)\, ds}\Phi(X_T)\right]$$

Proof: Let

$$Y_s = e^{\int_t^s r(u,X_u)\,du} V(s, X_s)$$

Then

$$\begin{aligned} dY_s &= r(s, X_s) e^{\int_t^s r(u,X_u)\,du} V(s, X_s)\,ds \\ &\quad + e^{\int_t^s r(u,X_u)\,du} \left[\nabla_x V \cdot \sigma\,dW_s + \left(\frac{\partial V}{\partial t} + \sum_{i=1}^n \mu_i \frac{\partial V}{\partial x^i} + \frac{1}{2} \sum_{i,j} C_{ij} \frac{\partial^2 V}{\partial x^i \partial x^j} \right) ds \right] \\ &= e^{\int_t^s r(u,X_u)\,du} \nabla_x V \cdot \sigma\,dW_s \end{aligned}$$

Thus $Y_T = Y_t + \int_t^T e^{\int_t^s r(u,X_u)\,du} \nabla_x V \cdot \sigma\,dW_s$. Now note that $Y_t = V(t, x)$, and that $Y_T = e^{\int_t^T r(s,X_s)\,ds} \Phi(X_T)$. Taking expectations yields the result.

□

Remarks 6.3.4 Consider a European contingent claim C on a share S with payoff $\Phi(S_T)$ at time T. Assuming that share prices follow a geometric Brownian motion (with constant drift and volatility), and that the interest rate r is constant, the Black–Scholes PDE to be solved is

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2} - rC = 0$$
$$C(T, S_T) = \Phi(S_T)$$

Using the Feynman–Kac Theorem, we obtain the solution as follows: Find the solution to the following SDE:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t$$

This diffusion is *not* the share price, although we've denote it by the same symbol. It is simply a solution to the above SDE, which is obtained directly from the Feynman–Kac formula. However, it looks exactly like the SDE for the share price in the *riskneutral world*!

Run it until time T. Then

$$\begin{aligned} C(0, S_0) &= \mathbb{E}^{0,S_0}\left[e^{\int_0^T -r\,dt} \Phi(S_T) \right] \\ &= e^{-rT} \mathbb{E}^{0,S_0}[\Phi(S_T)] \end{aligned}$$

Thus the price of the option is its discounted expected value, where the expectation is taken under a measure where S follows a geometric Brownian motion with drift rate r and variance rate σ.

□

> We have therefore reconciled the stochastic (riskneutral) and PDE approaches to pricing derivatives via the Feynman–Kac formula.

However, it should be pointed out that the riskneutral approach works in a general "semimartingale context" (where prices, rates, etc. are semimartingales), and not just in a "diffusion context" (where prices, rates, etc. are given as solutions to Itô diffusions, and are thus necessarily Markov processes). Hence the stochastic approach to finance is considerably more general.

Chapter 7

Financial Modelling in Continuous Time

7.1 Stochastic Financial Modelling

7.1.1 Basic Notions

We give here a quick introduction to the basics of stochastic financial modelling. We start with the following set–up:

- A market model is a tuple

  $$\mathcal{M} = (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}, (S_t^0, \ldots, S_t^N)_{t\geq 0})$$

  where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $(\mathcal{F}_t)_t$ a filtration satisfying the usaul conditions, and $S_t = (S_t^0, \ldots S_t^N)$ an $(N+1)$–dimensional adapted càdlàg semimartingale.

- We will often assume a finite horizon $[0, T]$, e.g. to price European options.

- We also make the usual assumptions about the market:
  - No transaction costs
  - Continuous trading
  - Liquid markets for every security
  - Short sales allowed
  - Perfect divisibility of assets

To get results, we will usually specialize: We will generally assume that Ω comes with a K–dimensional Brownian motion $W_t = (W_t^1, \ldots, W_t^K)$ which generates the filtration $\mathcal{F}_t$ (augmented to satisfy rthe usual conditions). We say that we have K *sources of noise*. Further, we assume that the asset price process is given by an Itô diffusion:

$$dS_t = \mu(t, S_t)\ dt + \sigma(t, S_t)\ dW_t$$

which is shorthand for

$$d\begin{pmatrix} S_t^0 \\ \vdots \\ S_t^N \end{pmatrix} = \begin{pmatrix} \mu^0(t, S_t) \\ \vdots \\ \mu^N(t, S_t) \end{pmatrix}\ dt + \begin{pmatrix} \sigma_{01}(t, S_t) & \ldots & \sigma_{0K}(t, S_t) \\ \vdots & \vdots & \vdots \\ \sigma_{N1}(t, S_t) & \ldots & \sigma_{NK}(t, S_t) \end{pmatrix}\ \begin{pmatrix} dW_t^1 \\ \vdots \\ dW_t^K \end{pmatrix}$$

Under these conditions, the asset price process is (strong) Markov.

Generally, we make another assumption on S_t^0: we assume that it is the money market account process ("riskless" bank account process), which has dynamics

$$dS_t^0 = rS_t^0 \, dt \qquad S_0^0 = 1$$

A *numéraire* is a price process N_t which has $N_t > 0$ a.s. Think of a numéraire as a *unit* into which other assets are translated. Thus if S_t is the price of S in money, then $\hat{S}_t = \frac{S_t}{N_t}$ is the price of S in units of N.

We often choose the numéraire to be the money market account process S_t^0. In that case, we write $\bar{S}_t = \frac{S_t}{S_t^0}$ for the value of S_t in terms of the numéraire. Of course, $\bar{S}_t$ is just the discounted value of S at time t.

A European contingent claim C is an derivative which, at some future time T has a payoff which is a known function of asset prices at time T, i.e.

$$C_T = f(S_T)$$

so that C_T is an $\mathcal{F}_T$–measurable random variable. The time T is called the *maturity* or *exercise time* of the claim.

A central problem is the pricing and hedging of such derivatives. A European claim can be priced by arbitrage methods only if there is a trading strategy which exactly replicates its payoff.

Definition 7.1.1

- A trading strategy/portfolio is a left–continuous (or, more generally, predictable) process $\phi_t = (\phi_t^0, \ldots, \phi_t^N)$ which is integrable w.r.t. the semimartingale S_t. ϕ_t^n is to be thought of as the number of asset S^n held in the portfolio at time t.

- The *value* process of the portfolio is given by

$$V_t(\phi) = \phi_t \cdot S_t = \sum_{n=0}^{N} \phi_t^n S_t^n$$

  and the *gains process* by

$$G_t(\phi) = V_t(\phi) - V_0(\phi)$$

- If N is a numéraire, we may also introduce num'eraire–deflated ("discounted") value– and gains processes by

$$\hat{V}_t(\phi) := \frac{V_t(\phi)}{N_t} \qquad \hat{G}_t(\phi) := \hat{V}_t(\phi) - \hat{V}_0(\phi)$$

- A trading strategy ϕ is self–financing if and only if $d(\phi_t \cdot S_t) = \phi_t \cdot dS_t$, i.e. if and only if

$$G_t(\phi) = \int_0^t \phi_u \cdot dS_u$$

□

Remarks 7.1.2
- The intuition behind the self–financing condition is a bit convoluted. Discretize time, and suppose that ϕ_t is the portfolio held over a period $[t, t+\Delta t]$. To be self–financing means that the value of the portfolio doesn't change purely because of rebalancing. Thus, at time t, the portfolio $\phi_{t-\Delta t}$ is rebalanced to become the portfolio ϕ_t. The value at time t of these portfolios is the same: $\phi_{t-\Delta t} \cdot S_t = \phi_t \cdot S_t$, i.e. $(\phi_t - \phi_{t-\Delta t}) \cdot S_t = 0$. It is tempting to deduce that, in the limit $\Delta t \to 0$, we obtain the self–financing condition
$$S_t \, d\phi_t = 0 \tag{$*$}$$
However, it would be wrong to use $(*)$ as the self–financing condition in continuous–time, because:

  (i) Stochastic integrals are to be interpreted in the Itô sense.

  (ii) If H_t is left–continuous, then the stochastic integral
  $$\int_0^T H_t \, dX_t = \lim_{||P|| \to 0} \sum H_{t_{n-1}}(X_{t_n} - X_{t_{n-1}})$$
  is a limit (in probability) of *left–hand* Riemann–Stieltjes sums.

  (iii) $S_t(\phi_t - \phi_{t-\Delta t})$ looks like a term in a *right–hand* sum.

  This problem is fixed rather easily: Add and subtract $S_{t-\Delta t}\Delta_t\phi$ from the left–hand side of $(*)$ to obtain:
  $$S_{t-\Delta t}(\phi_t - \phi_{t-\Delta t}) + (S_t - S_{t-\Delta t})(\phi_t - \phi_{t-\Delta t}) = 0$$
  In the continuous–time limit, this looks like
  $$S_t \, d\phi_t + d[S, \phi]_t = 0 \tag{$**$}$$
  because the first term *is* a left–hand sum, and the second term looks like a summand in the covariation process. By Itô's formula, we have $d(\phi_t \cdot S_t) = \phi_t \cdot dS_t + S_t \cdot d\phi_t + d[S, \phi]_t = \phi_t \cdot dS_t$. (But we must stress that the above argument is a purely intuitive formulation of the self–financing condition, as trading strategies need not be semimartingales, so that terms like $S_t \cdot d\phi_t$ and $d[S, \phi]_t$ need not make sense.)

- In the literature, other conditions are often imposed on trading strategies to ensure that they are sufficiently well–behaved. For example, a self–financing trading strategy is called *tame* if $V_t(\phi) \geq 0$ a.s. It is called *admissible* if its discounted value is a martingale under the EMM. This is important, because even the Black–Scholes model has "doubling" strategies, and is not arbitrage–free if arbitrarily large losses can be sustained. However, we will ignore these technical points in what follows.

□

Definition 7.1.3 A (European) contingent claim C is said to *attainable* if and only if there exists a self–financing strategy ϕ_t such that $C_T = V_T(\phi)$ (where T is the exercise date of the claim). Then ϕ is called a *replicating portfolio* for C.

A market model is *complete* if and only if every contingent claim is attainable.

□

Proposition 7.1.4 (Numéraire)
A self–financing portfolio remains self–financing under a change of numéraire.

From one point of view, this seems totally obvious: After all if we don't add or subtract funds from our portfolio when we reckon in units of money, we don't add or subtract funds if we reckon in units of barrels of oil either. However, our definition of self–financing is that $d(\phi_t \cdot S_t) = \phi_t \cdot \ dS_t$. Now suppose that we reckon in terms of a new numéraire N_t. Let $\hat{S}_t := \frac{S_t}{N_t}$ be the price of S in units of N. To prove that the self–financing condition holds, we must show that $d(\phi_t \cdot \hat{S}_t) = \phi_t \cdot \ d\hat{S}_t$ (i.e. that $\hat{G}_t(\phi) = \int_0^t \phi_u \cdot d\hat{S}_u$), and this no longer seems so obvious.

Proof: Let $\hat{V}_t = \frac{V_t(\phi)}{N_t}$. Then by Itô's formula

$$\begin{aligned} d\hat{V}_t &= \frac{1}{N_t}\, dV_t + V_t\, d\left(\frac{1}{N_t}\right) + d[V, \frac{1}{N}]_t \\ &= \frac{\phi_t}{N_t} \cdot dS_t + \phi_t S_t\, d\left(\frac{1}{N_t}\right) + \phi_t \cdot d[S, \frac{1}{N}] \end{aligned}$$

because (by the self–financing condition) $dV_t = \phi_t \cdot \ dS_t$, so $d[V, \frac{1}{N}]_t = \phi_t \cdot d[S, \frac{1}{N}]_t$. Thus

$$\begin{aligned} d\hat{V}_t &= \phi_t \cdot \left(\frac{1}{N_t}\, dS_t + S_t\, d\left(\frac{1}{N_t}\right) + d[S, \frac{1}{N}]\right) \\ &= \phi_t\, d\left(\frac{S_t}{N_t}\right) \\ &= \phi_t d\hat{S}_t \end{aligned}$$

□

Corollary 7.1.5 *If a contingent claim is attainable in a given numéraire, it is also attainable in any other numéraire, and the replicating portfolio is the same, i.e. if*

$$X = V_0 + \int_0^T \phi_t \cdot dS_t \qquad \textit{then} \qquad \hat{X} = \hat{V}_0 + \int_0^T \phi_t \cdot d\hat{S}_t$$

□

Proof: Suppose that ϕ is a self–financing strategy that replicates X, so that $X = V_T(\phi) = V_0 + G_T(\phi) = V_0 + \int_0^T \phi_t \cdot dS_t$. Since we have shown above that $d(\phi \cdot \hat{S}_t) = \phi_t \cdot d\hat{S}_t$, we obtain

$$\hat{X} = \hat{V}_T(\phi) = \hat{V}_0 + \hat{G}_T(\phi) = \hat{V}_0 + \int_0^T \phi_t \cdot d\hat{S}_t$$

□

In particular, if the numéraire is the bank account, then

$$\bar{V}_t(\phi) = \bar{V}_0(\phi) + \int_0^t \phi_u\, d\bar{S}_u$$

Remarks 7.1.6 A self–financing portfolio $\phi = (\phi^0, \dots, \phi^N)$ is completely determined by the N of the $N+1$ components. Thus, e.g., if we take the bank account to be the numéraire, and if we are given the risky asset components $\phi^1, \dots, \phi^N$, the value of the riskless asset component ϕ^0 is completely determined by the self–financing condition: Take S^0 to be the numéraire, so that

$$\phi_t \cdot \bar{S}_t = \bar{V}_t(\phi) = \bar{V}_0(\phi) + \sum_{n=0}^{N} \int_0^t \phi_u^n \, d\bar{S}_u^n = \bar{V}_0(\phi) + \sum_{n=1}^{N} \int_0^t \phi_u^n \, d\bar{S}_u^n$$

because $d\bar{S}_t^0 = 0$ — $\bar{S}_t^0 = 1$ is constant. Hence

$$\phi_t^0 = \bar{V}_0(\phi) + \sum_{n=1}^{N} \left[\int_0^t \phi_u^n \, d\bar{S}_u^n - \phi_t^n \bar{S}_t^n \right]$$

This is important, because it means that, given a portfolio of risky securities $(\phi_t^1, \dot{;} \phi_t^n)$, we can make it portfolio self–financing simply by adjusting the bank account. This will not affect the discounted gain of the portfolio at all, as $d\bar{S^0}_t = 0$. The same goes if we take S^0 to be a numéraire other than the bank account.

$\square$

7.1.2 Martingale Pricing

Recall that an arbitrage strategy is a trading strategy ϕ with the properties that

- $V_0(\phi) = 0$ — initial cost is zero.
- $\mathbb{P}(G_T \geq 0) = 1$, i.e. zero probability of a *loss*.
- $\mathbb{P}(G_T > 0) > 0$ — positive probability of a profit.

Exercise 7.1.7 It is often convenient to use a slightly different definition: Let N be a numéraire. Show that there is arbitrage if and only if there is a portfolio ϕ such that $\mathbb{P}(\hat{G}_T(\phi) \geq 0) = 1$ and $\mathbb{P}(\hat{G}_T(\phi) > 0) > 0$.

$\square$

Definition 7.1.8 *Suppose that N is a numéraire. A measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is an* equivalent martingale measure *(EMM) for numéraire N if an only if*

(i) $\mathbb{Q} \sim \mathbb{P}$;

(ii) $\hat{S} = (\frac{S_t}{N_t})_t$ *is a (local) $\mathbb{Q}$–martingale.*

If $\hat{S}_t$ is a $\mathbb{Q}$–martingale, $\mathbb{Q}$ is called a strong *EMM.*
An EMM associated with the money market account is called a riskneutral *measure.*

$\square$

If N is a numéraire, define $\hat{V}_t(\phi) = \frac{V_t(\phi)}{N_t}$, and define

$$\hat{G}_t(\phi) = \int_0^t \phi_u \, d\hat{S}_u$$

Note that if $\mathbb{Q}$ is an EMM for N, then both $\hat{V}$ and $\hat{G}$ are $\mathbb{Q}$–local martingales. Indeed, $\hat{G}_t = \int_0^t \phi_u \, d\hat{S}_u$ is a sum of stochastic integrals w.r.t. a $\mathbb{Q}$–local martingale.

We require $\mathbb{Q}$ to be equivalent to $\mathbb{P}$ so that both measures have the same arbitrage strategies: $\mathbb{P}(G_T > 0) > 0$ if and only if $\mathbb{Q}(G_T > 0) > 0$.

Further note that (ignoring some technical conditions):

- An arbitrage opportunity remains an arbitrage under
  - a change of equivalent measure;
  - a change of numéraire.
- A replicating portfolio remains a replicating portfolio under
  - a change of equivalent measure;
  - a change of numéraire.

Theorem 7.1.9 *If an EMM $\mathbb{Q}$ exists (for some numéraire N), then there are no arbitrage opportunities.*

Proof: If ϕ is a self–financing strategy, then

$$0 = \hat{G}_0(\phi) = \mathbb{E}_{\mathbb{Q}}[\hat{G}_T(\phi)]$$

Now because $\mathbb{P}(\hat{G}_T > 0) > 0$ if and only if $\mathbb{Q}(\hat{G}_T > 0) > 0$, and because $G_T \geq 0$ if and only if $\hat{G}_T \geq 0$, we cannot have both $G_T \geq 0$ and $\mathbb{E}_{\mathbb{P}}[G_T] > 0$. Thus ϕ cannot be an arbitrage, i.e. there are no arbitrage opportunities.

□

Example 7.1.10 The most common choice of numéraire is the money market account. Suppose that S_t^0 is the MMA, with price dynamics

$$dS_t^0 = r(t, \omega) S_t^0 \, dt$$

If $\mathbb{Q}$ is the EMM associated with S^0, then each $\bar{S}_t^n$ is a $\mathbb{Q}$– local martingale. Now

$$d\bar{S}_t^n = \frac{dS_t^n}{S_t^0} - \frac{S_t^n}{(S_t^0)^2} dS_t^0$$

and so

$$dS_t^n = S_t^0 d\bar{S}_t^n + r S_t^n \, dt = r S_t^n \, dt + dM_t^n$$

where $M_t^n = \int_0^t S_t^0 \, d\bar{S}_t^n$ is a $\mathbb{Q}$–local martingale. Conversely, if each $dS_t^n = r S_t^n \, dt + dM_t^n$ for some $\mathbb{Q}$–local martingale M_t^n, then $\mathbb{Q}$ is a riskneutral measure.

□

Theorem 7.1.11 (Martingale Valuation)
Suppose that X is an attainable contingent claim, and that $\mathbb{Q}$ is an EMM for numéraire N. Then

$$\hat{X}_t = \mathbb{E}_{\mathbb{Q}}[\hat{X}_T|\mathcal{F}_t]$$

i.e.

$$X_t = N_t \mathbb{E}_{\mathbb{Q}}\left[\frac{X_T}{N_T}|\mathcal{F}_t\right]$$

Proof: If ϕ replicates X, it does so under any numéraire, any EMM. Now by the Law of One Price, if $X_T = V_T(\phi)$, then $X_t = V_t(\phi)$ for all $t \leq T$. (Else buy the cheaper of the two and short the more expensive one at time t, and pocket the difference. At time T your gains will exactly match your obligations.) Thus

$$\hat{X}_t = \hat{V}_t(\phi) = \mathbb{E}_{\mathbb{Q}}[\hat{V}_T(\phi)|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\hat{X}_T|\mathcal{F}_t]$$

□

7.2 The Generalized Black–Scholes Model

As usual, we work in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that all information is contained in a filtration $(\mathcal{F}_t)_t$, which is generated by a K–dimensional standard $\mathbb{P}$–Brownian motion $W_t = (W_t^1, \ldots, W_t^K)$, augmented to satisfy the usual conditions. We further assume that there are N risky assets whose price processes $S_t = (S_t^1, \ldots S_t^N)$ are continuous semimartingales, indeed Ito diffusions, with dynamics of the form

$$dS_t^n = \mu_n(t) S_t^n \; dt + S_t^n \sum_{k=1}^{K} \sigma_{nk}(t) \; dW_t^k$$

or

$$\begin{pmatrix} dS_t^1 \\ \vdots \\ dS_t^N \end{pmatrix} = \begin{pmatrix} S_t^1 & 0 & \ldots & 0 \\ 0 & S_t^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & S_t^N \end{pmatrix} \left[\begin{pmatrix} \mu_1(t) \\ \vdots \\ \mu_N(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(t) & \ldots & \sigma_{1K}(t) \\ \vdots & \vdots & \vdots \\ \sigma_{N1}(t) & \ldots & \sigma_{NK}(t) \end{pmatrix} \begin{pmatrix} dW_t^1 \\ \vdots \\ dW_t^K \end{pmatrix} \right]$$

i.e

$$dS_t = D[S_t]\mu(t) \; dt + D[S_t]\sigma \; dW_t$$

where μ is a the *drift vector*, σ the *volatility matrix*, and $D[S_t]$ the diagonal matrix with asset prices along the diagonal.

We further assume that we have at our disposal a money market account A_t with dynamics

$$dA_t = r(t) A_t \; dt \qquad A_0 = 1$$

The money market account (MMA) will serve as our numéraire.

7.2.1 Construction of a Risk–Neutral Measure via Girsanov's Theorem

Let T be the horizon, i.e. we are only interested in the time interval $[0, T]$. Suppose that $\mathbb{Q}$ is an EMM for the MMA, so that $\bar{S}_t = \frac{S_t}{A_t}$ is an N–dimensional $\mathbb{Q}$–martingale. This means that the drift of each risky asset must be r under $\mathbb{Q}$, i.e. when we change the measure form $\mathbb{P}$ to $\mathbb{Q}$, the drift of S^n must change from μ_n to r.

To accomplish this change of measure via a Girsanov transformation, we need to find a kernel $\lambda(t) \in \mathbb{R}^k$ such that $\sigma(t)\lambda(t) = r(t) - \mu(t)$ (where r now doubles as the column vector whose entries are r). If we can find such a λ, then Girsanov's Theorem tells us that we can construct $\mathbb{Q}$ as follows: Let

$$\xi = \mathcal{E}(\int \lambda \cdot \, dW_t)_T = e^{\int_0^T \lambda(t)\cdot \, dW_t - \frac{1}{2}\int_0^T ||\lambda(t)||^2 \, dt}$$

and define $\mathbb{Q}$ by

$$\mathbb{Q}(F) = \int_F \xi \, d\mathbb{P}$$

Then

$$\hat{W}_t = W_t - \int_0^t \lambda(u) \, du$$

is a K–dimensional standard $\mathbb{Q}$–Brownian motion.

The dynamics of S_t under $\mathbb{Q}$ will therefore be

$$\begin{aligned} dS_t &= D[S_t](\mu(t, S_t) \, dt + \sigma(t, S_t)(d\hat{W}_t + \lambda \, dt)) \\ &= D[S_t](r(t) \, dt + \sigma(t, S_t) \, d\hat{W}_t) \end{aligned}$$

Hence the drift of each asset is indeed r under $\mathbb{Q}$, so that $\mathbb{Q}$ is a riskneutral measure.

Thus in order to construct a riskneutral measure, it is necessary that we are able to solve

$$\sigma\lambda = r - \mu$$

for λ. The above is a system of N linear equations in K unknowns, and will generally have a solution if $N \leq K$. If $N > K$, then the system is overdetermined, and a solution will only exist in special circumstances. However, if a solution λ does not exist, then we are unable to construct a riskneutral measure, and this means that there is arbitrage (see the next subsection). It follows that the force of arbitrage in the market will force those special circumstances to hold.

Example 7.2.1 Suppose we have a Black–Scholes model with two risky assets but only one source of noise, i.e. $N = 2$, $K = 1$:

$$\begin{aligned} dS_t^1 &= \mu_1 S_t^1 \, dt + \sigma_1 S_t^1 \, dW_t \\ dS_t^2 &= \mu_2 S_t^2 \, dt + \sigma_1 S_t^2 \, dW_t \end{aligned}$$

To find a riskneutral measure, we seek a Girsanov kernel λ solving

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} \lambda = \begin{pmatrix} r - \mu_1 \\ r - \mu_2 \end{pmatrix}$$

Here λ is a *number* (because $K = 1$). A solution will only exist if

$$\frac{r - \mu_1}{\sigma_1} = \frac{r - \mu_2}{\sigma_2}$$

as you will easily verify.

□

Remarks 7.2.2 The Girsanov kernel λ is very closely related to a quantity called the *market price of risk*. Consider a Black–Scholes model with only one source of noise, as in the example above. The we can only solve the system $\sigma\lambda = r - \mu$ if

$$\lambda = \frac{r - \mu_n}{\sigma_n} \quad \text{for all } n$$

The negative of this quantity,

$$-\lambda = \frac{\mu - r}{\sigma}$$

is called the *market price* of risk. We therefore see that, for there to be no arbitrage, all assets must have the same market price of risk.

The reason for the name *market price of risk* is as follows: $\mu - r$ is the excess rate of return of the asset (above the risk–free rate). Thus the ratio $u = \frac{\mu - r}{\sigma}$ can be interpreted as the *excess rate of return pr unit of volatility.*

In the case that we have more than one source of noise, each source can be ascribed its own market price of risk: The MPR of noise source W^k is simply $-\lambda_k$, the negative of the k^{th} component of the Girsanov kernel. We have

$$\sigma_{n1}(-\lambda_1) + \ldots \sigma_{nK}(-\lambda_K) = \mu_n - r$$

so a slight increase ε inthe volatility σ_{nk} corresponding to noise source W^k will result in an increase $(-\lambda_k)\varepsilon$ in the excess rate of return. Thus $-\lambda_k$ can be regarded as the excess rate of return caused by a unit change in the volatility coresponding to the k^{th} source of noise.

□

7.2.2 No–Arbitrage and the Existence of a Risk–Neutral Measure

We will sometimes denote the dot product of two vectors x, y by

$$x \cdot y = x^{tr} y = \langle x, y \rangle$$

where x^{tr} denotes the transpose of x. Observe that if A is a matrix, then $\langle x, Ay \rangle = \langle A^{tr} x, y \rangle$ (because $(A^{tr} y)^{tr} x = y^{tr} A x$). We recall here a lemma from linear algebra:

Lemma 7.2.3 *If σ is an $n \times d$–matrix (i.e. a linear operator $\sigma : \mathbb{R}^d \to \mathbb{R}^n$), then*

$$(\ker \sigma)^{\perp} = \text{ran } (\sigma^{tr})$$

Proof: $x \in \ker \sigma \Rightarrow \langle y, \sigma x \rangle = 0 \Rightarrow \langle \sigma^{tr} y, x \rangle = 0$ for all $y \in \mathbb{R}^n$. So $\sigma^{tr} y \perp \ker \sigma$ for all $y \in \mathbb{R}^n$, i.e. $\text{ran}(\sigma^{tr}) \subseteq (\ker \sigma)^{\perp}$. As $\dim \text{ran}(\sigma) + \dim \ker(\sigma) = d = \dim (\ker \sigma)^{\perp} + \dim \ker \sigma$ we see that $\dim \text{ran}(\sigma^{tr}) = \dim (\ker \sigma)^{\perp}$. Hence $\text{ran}(\sigma^{tr}) = (\ker \sigma)^{\perp}$.

⊣

Lemma 7.2.4 *If $P : \mathbb{R}^n \to \mathbb{R}^n$ is the orthogonal projection onto a subspace $V \subseteq \mathbb{R}^n$, then $\langle Px, y \rangle = \langle Px, Py \rangle = \langle x, Py \rangle$.*

Proof: If $y = Py + y^{\perp}$ is the orthogonal decomposition, then $\langle Px, y\rangle = \langle Px, Py + y^{\perp}\rangle = \langle Px, Py\rangle$.

$\dashv$

Suppose now that $dS_t = D[S_t](\mu_t\ dt + \sigma_t\ dW_t]$, where the asset price process S is N–dimensional, and W is a K–dimensional standard Brownian motion. The volatility process σ is a $N \times K$–matrix process. Let $P : \mathbb{R}^N \to \mathbb{R}^N$ be the orthogonal projection onto the subspace $\ker \sigma^{tr}$. Define

$$p(t) := P(\mu(t) - r\mathbf{1})$$

We omit the (technical) proof that $p(t)$ is measurable. Further, for $n = 1, \ldots, N$ define portfolio components

$$\theta^n(t) = \begin{cases} \dfrac{p(t)}{||p(t)||\bar{S}_t^n} & \text{if } p(t) \neq 0 \\ 0 & \text{else} \end{cases}$$

and choose θ^0 to make the portfolio self–financing with inital value 0. Then

$$\begin{aligned} \bar{G}_T(\theta) &= \int_0^T \theta_t \cdot\ d\bar{S}_t \\ &= \int_0^T I_{\{p(t)\neq 0\}} \frac{p(t)\cdot(\mu - r\mathbf{1})}{||p(t)||}\ dt + \int_0^T \frac{p(t)\cdot\sigma}{||p(t)||} I_{\{p(t)\neq 0\}}\ dW_t \\ &= \int_0^T I_{\{p(t)\neq 0\}} \frac{||p(t)||^2}{||p(t)||}\ dt \\ &= \int_0^T I_{\{p(t)\neq 0\}} ||p(t)||\ dt \end{aligned}$$

because $p(t)\cdot(\mu - r\mathbf{1}) = p(t)\cdot p(t) = ||p(t)^2||$, and because $\sigma \cdot p(t) = 0$, as $p(t) \in \ker(\sigma^{tr})$.

Thus $\bar{G}_T(\theta) > 0$ a.s. unless $p(t) = 0$ a.e., i.e. unless $\mu - r\mathbf{1} \in (\ker\sigma^{tr})^{\perp} = \operatorname{ran}\sigma$. Now if there is no arbitrage, then we cannot have $\bar{G}_T(\theta) > 0$, and so there must be λ such that $\sigma\lambda = \mu - r\mathbf{1}$.

Given such a λ, the previous subsection shows that $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(-\lambda \bullet W)_T$ defines a risk–neutral measure. We have therefore proved:

Theorem 7.2.5 (Fundamental Theorem of Asset Pricing for the Generalized Black–Scholes Model)
The Generalized Black–Scholes Model is arbitrage–free iff and only if there is a risk–neutral measure.

7.2.3 Hedging of European Contingent Claims

We can use Thm. 7.1.11 to price a European–style contingent claim X *only* when we know that X is attainable. To find a replicating portfolio θ so that $V_T(\theta) = X_T$ may be very difficult, however. Provided that — roughly speaking — the underlying assets contain all the information, the Martingale Representation Theorem may overcome this difficulty, by showing that *all* contingent claims are attainable.

Suppose we have an arbitrage–free SSM $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0\leq t\leq T}, (S_t^0, \ldots, S_t^N)_{0\leq t\leq T})$. We assume that all information is contained in the filtration $(\mathcal{F}_t)_t$ is generated by a K–dimensional

standard $\mathbb{P}$–Brownian motion $W_t = (W_t^1, \ldots, W_t^K)$, (augmented to satisfy the usual conditions). We further assume that the N risky assets whose price processes $S_t = (S_t^1, \ldots S_t^N)$ are Itô diffusions, with dynamics of the form

$$dS_t = D[S_t]\mu_t \, dt + D[S_t]\sigma_t \, dW_t$$

where μ is a the *drift vector*, σ the *volatility matrix*, and $D[S_t]$ the diagonal matrix with asset prices along the diagonal. Because the model is assumed to be arbitrage–free, we have at our disposal at least one risk–neutral measure $\mathbb{Q}$, which is obtained from $\mathbb{P}$ via a Girsanov transformation with kernel some K–dimensional predictable process λ. The process $W_t^{\mathbb{Q}} := W_t - \int_0^t \lambda_s \, ds$ is a $\mathbb{Q}$–Brownian motion. The discounted asset dynamics under $\mathbb{Q}$ are thus

$$d\bar{S}_t = D[\bar{S_T}]\sigma_t \, dW_t^{\mathbb{Q}}$$

Suppose also that we have a european–style contingent claim X, with payoff X_T at expiry T. Define a $(\mathbb{Q}, (\mathcal{F}_t)_t)$–martingale $(M_t)_{t \leq T}$ by

$$M_t := \mathbb{E}_{\mathbb{Q}}[\bar{X}_T | \mathcal{F}_t]$$

By the martingale representation theorem, there is a K–dimensional predictable process H_t so that

$$\bar{X}_T = M_T = \mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T H_t \cdot \, dW_t^{\mathbb{Q}}$$

But $d\bar{S}_t = D[\bar{S_T}]\sigma_t \, dW_t^{\mathbb{Q}}$. If we can find a left–inverse $\hat{\sigma}$ for the matrix σ, we would get

$$dW_t^{\mathbb{Q}} = \hat{\sigma}_t D[\bar{S}_t]^{-1} \, d\bar{S}_t$$

Now put

$$H_t \hat{\sigma}_t D[\bar{S}_t]^{-1}$$

be the risky–asset component of a portfolio θ, i.e.

$$\theta_t^n = \frac{1}{\bar{S}_t^n} \sum_{k=1}^K \hat{\sigma}_{kn}(t) H_t^k \qquad n = 1, \ldots, N$$

and choose θ^0 to make the portfolio self–financing with initial value $\mathbb{E}_{\mathbb{Q}}[\bar{X}_T]$. Then

$$\begin{aligned}
\bar{V}_T(\theta) &= \bar{V}_0(\theta) + \bar{G}_T(\theta) \\
&= \mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T \theta_t \, d\bar{S}_t \\
&= \mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T H_t \hat{\sigma}_t D[\bar{S}_t]^{-1} D[\bar{S}_t]\sigma_t \, dW_t^{\mathbb{Q}} \\
&= \mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T H_t \, dW_t^{\mathbb{Q}} \\
&= \bar{X}_T
\end{aligned}$$

Thus θ is a replicating portfolio for X.

It follows that if σ has a left inverse, then *any* contingent claim is attainable, i.e. the market is complete. We now give a rough, intuitive argument for when we will be able to replicate an arbitray X. In the above, we require a θ so that

$$\mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T H_t \cdot dW_t^{\mathbb{Q}} = \mathbb{E}_{\mathbb{Q}}[\bar{X}_T] + \int_0^T \theta_t \cdot d\bar{S}_t$$

so we require

$$H_t \cdot dW_t^{\mathbb{Q}} = \theta_t \cdot D[\bar{S}_t \sigma \cdot dW_t^{\mathbb{Q}}$$

This means that we must solve, at each instant, the linear system of equations

$$H = \theta \cdot D[\bar{S}]\sigma$$

for the risky components $\theta^1, \ldots, \theta^N$, i.e. we have a system of K equations in N unknowns. Roughly speaking, this means that we will have a solution as soon as there are more variables than constraints, i.e. when $N \geq K$. We expect the solution to be unique when $K = N$.

To find a risk–neutral measure, we must find a Girsanov kernel, i.e. a solution λ to the linear system $\sigma\lambda = r - \mu$. This is a system of N equations in K unknowns, and, roughly speaking, will have a solution when $K \geq N$, and a unique solution when $K = N$.

Thus: If $K < N$, then we expect to be able to hedge every contingent claim in more than one way. We therefore expect there to be arbitrage, as different replicating portfolios need not have the same value. Thus we don't expect a risk–neutral measure to exist. On the other hand, if $K > N$, we expect that there will be many risk–neutral measures, and there will be contingent claims that cannot be replicated. These unattainable contingent claims may have different prices under different risk–neutral measures.

The sweet spot is therefore $K = N$ — as many sources of noise W as assets S: We expect that a unique risk–neutral measure exists, and that every contingent claim has a unique replicating portfolio. For that, we require σ to be invertible.

7.2.4 European Vanilla Call and Put Options

The following formula is basic for "lognormal pricing" and underlies not only the Black–Scholes formula for call– and put options, but also Black–type formulas for futures options, Margrabe options, caps and floors, swaptions, etc.

Proposition 7.2.6 *If X is lognormal, with $\ln X \sim N(\mu, s^2)$, then*

$$\mathbb{E}[(X-K)^+] = \mathbb{E}[X]N(d_+) - KN(d_-)$$

where

$$d_\pm = \frac{\ln\frac{\mathbb{E}X}{K} \pm \frac{1}{2}s^2}{s} \qquad \text{and} \qquad \mathbb{E}X = e^{\mu + \frac{1}{2}s^2}$$

Proof: Let $Z := \frac{\ln X - \mu}{s}$ so that $X = e^{\mu + sZ}$ and $Z \sim N(0,1)$. Clearly $\mathbb{E}X = e^{\mu}\mathbb{E}[e^{sZ}] = e^{\mu + \frac{1}{2}s^2}$, using the moment generating function of a standard normal random variable. Then,

using the symmetry of the standard normal random distribution,

$$\begin{aligned}
\mathbb{E}[(X-K)^+] &= \frac{1}{\sqrt{2\pi}}\int_{\frac{\ln K-\mu}{s}}^{\infty}(e^{\mu+sz}-K)e^{-\frac{z^2}{2}}\,dz \\
&= \frac{e^{\mu+\frac{1}{2}s^2}}{\sqrt{2\pi}}\int_{\frac{\ln K-\mu}{s}}^{\infty} e^{-\frac{(z-s)^2}{2}}\,dz - \frac{K}{\sqrt{2\pi}}\int_{\frac{\ln K-\mu}{s}}^{\infty} e^{-\frac{z^2}{2}}\,dz \\
&= \mathbb{E}[X]\mathbb{P}(Z+s\geq \tfrac{\ln K-\mu}{s}) - K\mathbb{P}(Z\geq \tfrac{\ln K-\mu}{s}) \\
&= \mathbb{E}[X]\mathbb{P}(Z\leq \tfrac{\mu-\ln K}{s}+s) - K\mathbb{P}(Z\leq \tfrac{\mu-\ln K}{s}) \\
&= \mathbb{E}[X]\mathbb{P}\Big(Z\leq \frac{\ln\frac{\mathbb{E}[X]}{K}+\frac{1}{2}s^2}{s}\Big) - K\mathbb{P}\Big(Z\leq \frac{\ln\frac{\mathbb{E}[X]}{K}-\frac{1}{2}s^2}{s}\Big)
\end{aligned}$$

We prefer to use the expression $\mathbb{E}[X]$, rather than $e^{\mu+\frac{1}{2}s^2}$, in the above formula, because very often X will be the terminal value of some martingale, in which case $\mathbb{E}[X]$ is its initial value, and needs not be calculated as it is already known.

The Black–Scholes formula for vanilla call options now follows easily: Given (one–dimensional) risk–neutral asset dynamics

$$dS_t = S_t(r\,dt + \sigma\,dW_t)$$

with constant riskless rate r, we obtain discounted dynamics $d\bar{S}_t = \bar{S}_t\sigma\,dW_t$, so that $\bar{S}_T = S_0e^{-\frac{1}{2}\sigma^2T+\sigma W_T}$. Hence $\bar{S}_T$ is lognormal with $\ln\bar{S}_T \sim N(-\frac{1}{2}\sigma^2T, \sigma^2T)$ under the risk–neutral measure $\mathbb{Q}$. Also, since the discounted asset price process $\bar{S}_t$ is a martingale under the measure $\mathbb{Q}$, we have $\mathbb{E}_{\mathbb{Q}}[\bar{S}_T] = S_0$.

Now let C be a european call option with strike K and expiry T on the underlying asset S. By the above lognormal formula, and the risk–neutral valuation formula, the $t=0$–price of the call is given by

$$C_0 = \mathbb{E}_{\mathbb{Q}}[\overline{(S_T-K)^+}] = \mathbb{E}_{\mathbb{Q}}[(\bar{S}_T - Ke^{-rT})^+] = S_0N(d_+) - Ke^{-rT}N(d_-)$$

where

$$d_\pm = \frac{\ln\frac{\mathbb{E}_{\mathbb{Q}}[\bar{S}_T]}{Ke^{-rT}} \pm \frac{1}{2}\sigma^2T}{\sigma\sqrt{T}} = \frac{\ln\frac{S_0}{K} + (r\pm\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

The fact that $(S_T-K) = (S_T-K)^+ - (K-S_T)^+$ — so-called put–call parity — allows us to easily calculate the price for a european put option P:

$$P_0 = e^{-rT}\mathbb{E}_{\mathbb{Q}}[(K-S_T)^+] = e^{-rT}\mathbb{E}_{\mathbb{Q}}[(S_T-K)^+ - (S_T-K)] = C_0 - S_0 + Ke^{-rT}$$

Rearrangement and symmetry of the standard normal distribution now yield

$$P_0 = Ke^{-rT}N(-d_-) - S_0N(-d_+)$$

7.2.5 Caveat: Arbitrage in the Black–Scholes Model

Throughout, we have made certain simplifying assumptions to keep the technical machinery to a minimum. For example, we have assumed that local martingales are martingales, which may not be the case. We give here a *pathological* example of arbitrage in an "arbitrage–free" model. It is related to the doubling strategy in gambling: Bet 1 on the toss of a coin. If you lose, bet 2 on the next toss. If you lose again, bet 4 on the next toss. With probability 1 you

will eventually win. If this is on the $(N+1)^{th}$ toss, you will win 2^N, whereas your losses up to that time will be $1+2+4+\cdots+2^{N-1}$. Hence your total gain is $2^N-(1+2+\ldots 2^{N-1})=1$. Thus a single win suffices to recoup all previous losses, and you are guaranteed to win eventually.

In the strategy below, we increase the bet on the stock when the stock price goes down. Eventually, the stock price will move up, recouping all our previous losses.

Exercise 7.2.7 Arbitrage in a simple Black–Scholes Model
Consider a financial markets model $(\Omega,\mathcal{F},\mathbb{P},\mathcal{F}_t,(S_t,A_t))$ with a single risky share S (in addition to the riskless bank account A), where $r=\mu=0$ and $\sigma=1$, i.e. the dynamics are given by

$$dS_t = S_t\,dW_t \qquad dA_t = 0$$

where W_t is a one–dimensional standard Brownian motion. Note that $\mathbb{P}$ is a risk–neutral measure, because S_t, A_t are $\mathbb{P}$–local martingales. Define

$$I_t = \int_0^t \sqrt{\frac{1}{T-s}}\,dW_s$$

(a) Show that $[I]_t = \ln\frac{T}{T-t}$ for $t\in[0,T)$.

(b) Let $g(s)=T(1-e^{-s})$ for $s\in[0,\infty)$, and define $X_s=[I]_{g(s)}$. Use Lévy's characterization to show that $(X_t)_t$ is a Brownian motion.

(c) Deduce that $\limsup_{t\uparrow T} I(t)=\infty$ a.s. and $\liminf_{t\uparrow T} I(t)=-\infty$ a.s.

(d) Now let $\alpha>0$, and define a stopping time τ_α by

$$\tau_\alpha = \inf\{t : I_t=\alpha\}$$

Explain why $0<\tau_\alpha<T$ a.s.

(e) Define a portfolio $\varphi=(\varphi^A,\varphi^S)$ by

$$\varphi_t^S = \frac{1}{S_t\sqrt{T-t}} I_{t\le\tau_\alpha}$$

and adjust φ^A to ensure that φ is self–financing with initial value $V_0(\varphi)=0$. (Note that φ_t^S increases if S_t decreases, and also as $t\to T$.) Show that $V_T(\varphi)=\alpha$ a.s.

□

In the above exercise, we saw that there are arbitrage strategies in the Black–Scholes model. We now show that there are no *admissible* arbitrage strategies. A portfolio θ is said to be *admissible* if and only if there is a constant $C\le 0$ such that $V_t(\theta)\ge C$ for all $0\le t\le T$ — the portfolio may not fall into a debt which is $\ge |C|$. This is a realistic assumption, as your broker or creditors will close out your position if it becomes too negative.

Exercise 7.2.8 Let $\mathbb{Q}$ be a risk–neutral measure (so that $\bar{S}_t$ is a $\mathbb{Q}$–local martingale).

(a) Use Fatou's Lemma to show that a non–negative local martingale is a supermartingale.

(b) Suppose that φ is a self–financing trading strategy such that $V_0(\varphi)=0$ and $\bar{\bar{V}}_t(\varphi)\ge C$ for $0\le t\le T$ and a constant $C>-\infty$. Show that $V_t(\varphi)$ is a supermartingale.

(c) Now conclude that φ cannot be an arbitrage.

□

7.3 The Black–Scholes PDE

Consider a market model that is complete and arbitrage–free, so that there are as many sources of noise as there are risky securities. We want to price contingent claim C with payoff $\Phi(S_T)$ at expiry T. We assume that the price C_t of the contingent claim at some prior time t is given by a sufficiently smooth function $C_t = F(t, S_t) : \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$. For there to be a *fair* or *rational* price at all, we must assume that the market is arbitrage–free to begin with, and that the addition of the new security V does not introduce any arbitrage opportunities.

Assume that the risky asset dynamics are given by the usual multidimensional SDE $dS_t = D[S_t]\mu \; dt + D[S_t]\sigma \; dW_t$ (where $S_t = (S_t^1, \ldots, S_t^N)$ and$W$ is an N–dimensional Brownian motion), and that the MMA satisfies $dA_t = rA_t \; dt$.

Now form a portfolio V consisting of one derivative C, as well as a combination of risky assets and the MMA. The aim is to make the portfolio locally riskless. An arbitrage argument then shows that the portfolio must have the same return as the MMA, and this will allow us to derive a PDE.

Let $w_n(t)$ be the relative weight of asset S^n in the portfolio V, w_A the weight of the MMA, and W_C the weight of the contingent claim, so that

$$\frac{dV}{V} = \sum_n w_n \frac{dS^n}{S^n} + w_A \frac{dA}{A} + w_C \frac{dC}{C} \tag{*}$$

as you can easily verify. Using the asset dynamics above, as well as

$$\begin{aligned} dC &= \frac{\partial C}{\partial t} \; dt + \sum_{n=1}^N \frac{\partial C}{\partial S^n} \; dS^n + \frac{1}{2} \sum_{n,m} \frac{\partial^2 C}{\partial S^n \partial S^m} \; d[S^n, S^m] \\ &= \mu_C C \; dt + \sigma_C C \; dW_t \end{aligned}$$

where

$$\begin{aligned} \mu_C &= \frac{1}{C} \left[\frac{\partial C}{\partial t} + \sum_{n=1}^N \mu_n S^n \frac{\partial C}{\partial S^n} + \frac{1}{2} \sum_{n,m} (\sigma \sigma^{tr})_{nm} S^n S^m \frac{\partial^2 C}{\partial S^n \partial S^m} \right] \\ \sigma_C &= \frac{1}{C} \sum_{n=1}^N S^n (\sigma_{n1}, \ldots, \sigma_{nN}) \frac{\partial C}{\partial S^n} = \frac{1}{C} \sum_n S^n \sigma_n \frac{\partial C}{\partial S^n} \end{aligned}$$

where $\sigma_n = (\sigma_{n1}, \ldots, \sigma_{nN})$ is the n^{th} row of the volatility matrix.

Before we put this into $(*)$, note that

$$w_A = 1 - (w_C + \sum_m w_n)$$

because weights sum to 1. Plugging this into $(*)$, we obtain

$$\frac{dV}{V} = \left[\sum_{n=1}^N w_n(\mu_n - r) + w_C(\mu_C - r) + r \right] \; dt + \left[\sum_{n=1}^N w_n \sigma_n + w_C \sigma_C \right] \; dW_t$$

In order to make this portfolio locally riskless, we must choose the weights so that

$$\sum_{n=1}^N w_n \sigma_n + w_C \sigma_C = 0 \tag{**}$$

This is a system of N linear equations in the $N+1$ unknowns $w_1, \dots, w_N, w_C$.

Assuming that this system can be solved, we now have

$$\frac{dV}{V} = \left[\sum_{n=1}^{N} w_n(\mu_n - r) + w_C(\mu_C - r) + r \right] dt$$

Now let β be some positive constant, and let's attempt to get a riskless return of $r + \beta$ on the portfolio V, i.e.

$$\sum_{n=1}^{N} w_n(\mu_n - r) + w_C(\mu_C - r) + r = r + \beta$$

so that

$$\sum_{n=1}^{N} w_n(\mu_n - r) + w_C(\mu_C - r) = \beta$$

Combining this with $(**)$, we get a linear system

$$\begin{pmatrix} \mu_1 - r & \mu_2 - r & \dots & \mu_C - r \\ \sigma_1^{tr} & \sigma_2^{tr} & \dots & \sigma_C^{tr} \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ w_C \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

i.e.

$$H \begin{pmatrix} w_S \\ w_C \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix}$$

Thus to get our excess riskless return of β we need to solve a system of $N+1$ equations in $N+1$ unknowns. A solution will exist if and only if $\det(H) \neq 0$.

Now in an arbitrage–free market, it is impossible to obtain a riskless return above the risk–free rate. Thus if the market is arbitrage–free, the matrix H must be *singular*, i.e. be non–invertible, i.e. have zero determinant. The same is true for H^{tr}, the transpose of H. If we consider the transpose

$$H^{tr} = \begin{pmatrix} \mu_1 - r & \sigma_1 \\ \mu_2 - r & \sigma_2 \\ \vdots & \vdots \\ \mu_N - r & \sigma_N \\ \mu_C - r & \sigma_C \end{pmatrix}$$

then the singularity of H^{tr} implies that its columns are not linearly independent. Thus the first column of H^{tr} can be expressed as a linear combination of the other columns, i.e. there exists $u = (u_1, \dots, u_N)^{tr}$ such that

$$\sigma u = \mu - r \qquad \sigma_C u = \mu_C - r$$

wheer $\mu = (\mu_1, \dots, \mu_N)^{tr}$, etc. Clearly the u_n are simply the market prices of risk corresponding to the sources $W^1, \dots, W^n$, i.e. $-u$ is the Girsanov kernel effecting the change of measure from real world to riskneutral. Now is σ is invertible, we must have

$$u = \sigma^{-1}(\mu - r) \quad \text{and thus} \quad \mu_C - r = \sigma_C \sigma^{-1}(\mu - r)$$

But

$$\sigma_C = \frac{1}{C}\sum_{n=1}^{N} S^n(\sigma_{n1},\ldots,\sigma_{nN})\frac{\partial C}{\partial S^n} = \frac{1}{C}(S^1\frac{\partial C}{\partial S^1},\ldots,S^N\frac{\partial C}{\partial S^N})\cdot\sigma$$

and so

$$\mu_C - r = \frac{1}{C}(S^1\frac{\partial C}{\partial S^1},\ldots,S^N\frac{\partial C}{\partial S^N})\cdot(\mu - r)$$

Now

$$\mu_C = \frac{1}{C}\left[\frac{\partial C}{\partial t} + \sum_{n=1}^{N}\mu_n S^n\frac{\partial C}{\partial S^n} + \frac{1}{2}\sum_{n,m}(\sigma\sigma^{tr})_{nm}S^nS^m\frac{\partial^2 C}{\partial S^n\partial S^m}\right]$$

which yields the generalized Black–Scholes equation:

$$\boxed{\frac{\partial C}{\partial t} + \sum_n rS^n\frac{\partial C}{\partial S^n} + \frac{1}{2}\sum_{n,m}(\sigma\sigma^{tr})S^nS^m\frac{\partial^2 C}{\partial S^n\partial S^m} - rC = 0}$$

7.4 Correlated Brownian Motions

When many assets are available in the economy, it is unrealistic to assume that these are all driven by only one source of noise. It would be equally unrealistic, however, to assume that all are driven by separate, indpendent, Brownian motions. Thus it becomes necessary to generate multiple correlated Brownian motions.

Let's first consider a simpler case, where we are trying to generate *not* correlated Brownian motions, *just* correlated normal random variables, i.e. suppose that we want to generate mean zero normal random variables $X_1,\ldots,X_n$ with a specific covariance matrix $\Sigma = (\sigma_{ij})$. Here $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$.

You can check the following simple

Fact: If $(X_1,\ldots,X_n)$ is a random vector with covariance matrix Σ and if A is an $n\times n$–matrix, then the random vector

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

has covariance matrix $A\Sigma A^{tr}$.

Indeed, we may assume without loss of generality that $\mathbb{E}[\mathbf{X}] = \mathbf{0}$. Then $\mathrm{Cov}(\mathbf{Y}) = \mathbb{E}[\mathbf{Y}\mathbf{Y}^{tr}] = A\mathbb{E}[\mathbf{X}\mathbf{X}^{tr}]A^{tr}$.

□

Now recall that a matrix C is said to be positive semidefinite if and only if $\vec{x}^{tr}C\vec{x} \geq 0$ for all vectors $\vec{x}$. Covariance matrices are necessarily symmetric positive semidefinite, since if $\Sigma = \mathrm{Cov}(\mathbf{X})$, then $\vec{x}^{tr}\Sigma\vec{x} = \mathrm{Var}(\vec{x}^{tr}\mathbf{X})$ is the variance of the random variable $\vec{x}^{tr}\mathbf{X}$, and hence non–negative.

It is known that symmetric positive semidefinite matrices have a *Cholesky decomposition*, which means that it is possible to find a (real) lower triangular matrix C such that

$$\Sigma = CC^{tr}$$

Note that if C is an arbitrary matrix, then CC^{tr} is necessarily symmetric (obvious), and positive semidefinite: If $\vec{x}$ is a column vector, then $\vec{x}^{tr}C$ is a row vector, with length given by $||\vec{x}^{tr}C||^2 = (\vec{x}^{tr}C)(\vec{x}^{tr}C)^{tr} = \vec{x}^{tr}CC^{tr}\vec{x}$. Since the length of a vector is necessarily non–negative, CC^{tr} is positive semidefinite.

Thus any matrix A that can be written as $A = CC^{tr}$ is necessarily symmetric positive semidefinite. By the Cholesky decomposition, the reverse is also true. Indeed we can easily find a lower triangular C which does the trick. There is no deep mathematics behind this — we merely need to solve

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \begin{pmatrix} c_{11} & c_{21} & \dots & c_{n1} \\ 0 & c_{22} & \dots & c_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & c_{nn} \end{pmatrix}$$

This system is easily solved: $c_{11}^2 = a_{11}$ gives us c_{11}. $c_{11}c_{21} = a_{12}$ now gives us c_{21}, etc.

There are fast algorithms available for calculating Cholesky decompositions.

Now suppose we are able to generate independent standard normal random variables $Z_1, \dots, Z_n$. These have the identity matrix as covariance matrix. Define a random vector $\mathbf{X} = C\mathbf{Z}$. Then the covariance matrix of $\mathbf{Z}$ is simply $CIC^{tr} = CC^{tr} = \Sigma$. Thus to get a vector $\mathbf{X}$ of mean zero normally distributed random variables with covariance matrix Σ, proceed as follows:

- Generate a vector $\mathbf{Z}$ of independent standard normal random variables (of the same dimension as $\mathbf{X}$).
- Find the Cholesky decomposition $\Sigma = CC^{tr}$ of the symmetric positive semidefinite matrix Σ.
- Put $\mathbf{X} = C\mathbf{Z}$

Note that if $\Sigma = (\sigma_{ij})$ is a covariance matrix, then the correlation matrix is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Clearly the correlation matrix is also symmetric.

Now to obtain *correlated Brownian motions* $W_t^1, \dots, W_t^n$, we can proceed in a similar way. But first: What exactly do we *mean* if we say two Brownian motions W^1, W^2 are correlated? Clearly this has meaning if we speak about *changes* in the processes. If W^1, W^2 are highly correlated, then we expect a positive change in W^1 to be accompanied by a positive change in W^2.

Now suppose that we have *independent* standard Brownian motions $B_t^1, \dots, B_t^n$. Consider a matrix $\Gamma = (\gamma_{ij})$ with the property that all the rows of Γ have *unit length.* Define

$$\begin{pmatrix} W_t^1 \\ \vdots \\ W_t^n \end{pmatrix} = \Gamma \begin{pmatrix} B_t^1 \\ \vdots \\ B_t^n \end{pmatrix}$$

so that each $W_t^i = \sum_j \gamma_{ij} B_t^j$ is a linear combination of B_t^j's. It follows that each W_t^i is a continuous local martingale. Now

$$\begin{aligned}[W^i]_t &= \left(\sum_j \gamma_{ij}\right)\left(\sum_k \gamma_{ik}\right)[B^j, B^k]_t \\ &= \sum_j \gamma_{ij}^2 t \\ &= t\end{aligned}$$

because $[B^j, B^k]_t = \delta_{jk} t$, and $\sum_j \gamma_{ij}^2 = 1$ (being the square of the length of the i^{th} row of Γ).. Hence, by Lévy's Characterization, each W_t^i is a Brownian motion. Now

$$\begin{aligned}[W^i, W^j]_t &= \sum_{k,l} \gamma_{ik}\gamma_{jl}\delta_{kl} t \\ &= (\Gamma\Gamma^{tr})_{ij} t\end{aligned}$$

which we may also write as

$$dW_t^i \, dW_t^j = (\Gamma\Gamma^{tr})_{ij} \, dt$$

Hence

$$\mathbb{E}[W_t^i W_t^j] = \mathbb{E}[\,[W^i, W^j]_t] = (\Gamma\Gamma^{tr})_{ij} t$$

Thus W^i, W^j are correlated Brownian motions, and the correlation between W_t^i and W_t^j is simply $(\Gamma\Gamma^{tr})_{ij}$, independent of t (because the variance of each W_t^i is just t).

Note that if Σ, ρ are, respectively, the covariance and correlation matrix of $(W_t^1, \ldots, W_t^n)$, then $\Sigma = \rho t$. Hence ρ is also symmetric positive semidefinite, and thus has a Cholesky decomposition $\rho = \Gamma\Gamma^{tr}$.

Further note that not every symmetric positive semidefinite matrix can be the correlation matrix of some multidimensional Brownian motion: Since the correlation of a random variable with itself is 1, it is necessary that a correlation matrix has 1's down the diagonal. This, in turn, implies that the Cholesky decomposition matrix Γ will have row vectors of unit length.

Hence, to create correlated Brownian motions with correlation matrix ρ, proceed as follows:

- Find the Cholesky decomposition $\rho = \Gamma\Gamma^{tr}$. Γ will have rows of unit length.
- Define $\mathbf{W} = \Gamma\mathbf{B}$, where $\mathbf{B}$ is a multidimensional standard Brownian motion (with independent component processes). $\mathbf{W}$ will be a multidimensional Brownian motion with correlation matrix ρ.

One final remark about differential notation: Since $\Gamma\Gamma^{tr} = \rho$, and since $dW_t^i \, dW_t^j = (\Gamma\Gamma^{tr})_{ij} \, dt$, we have

$$d[W^i, W^j]_t = dW_t^i \, dW_t^j = \rho_{ij} \, dt$$

Example 7.4.1 To create two correlated Brownian motions W_t^1, W_t^2 with correlation ρ (a number, not a matrix), proceed as follows: The correlation matrix is

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Its Cholesky decomposition is found by solving

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \qquad (= \Gamma\Gamma^{tr})$$

for a, b, c. (Recall that Γ is lower triangular.) Thus $a = 1, b = \rho, c = \sqrt{1-\rho^2}$. Gratifyingly, the rows of Γ are seen to possess unit length.

Finally, if B_t^1, B_t^2 are standard independent Brownian motions, then

$$\begin{aligned} W_t^1 &= B_t^1 \\ W_t^2 &= \rho B_t^1 + \sqrt{1-\rho^2} B_t^2 \end{aligned}$$

are Brownian motions with correlation ρ.

□

Example 7.4.2 Suppose we have asset dynamics

$$\begin{pmatrix} dS_t^1 \\ dS_t^2 \end{pmatrix} = \begin{pmatrix} 0.3S_t^1 \\ 0.2S_t^2 \end{pmatrix} dt + \begin{pmatrix} 0.1S_t^1 & 0.4S_t^1 \\ 0.4S_t^2 & 0.3S_t^2 \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix}$$

where W_t^1, W_t^2 are independent Brownian motions. Here each asset is driven by two sources of noise. It may be convenient to rewrite the dynamics in a decoupled fashion:

$$\begin{aligned} dS_t^1 &= 0.3S_t^1 \, dt + \hat{\sigma}_1 S_t^1 \, d\hat{W}_t^1 \\ dS_t^2 &= 0.2S_t^2 \, dt + \hat{\sigma}_2 S_t^2 \, d\hat{W}_t^2 \end{aligned}$$

where $\hat{W}_t^1, \hat{W}_t^2$ are *correlated* Brownian motions. This *may* be simpler, because each asset is now driven by only one source of noise.

The two things that we need to know are:

(i) What are the volatilities $\hat{\sigma}_1, \hat{\sigma}_2$?

(ii) What is the correlation ρ between $\hat{W}_t^1$ and $\hat{W}_t^2$?

Clearly, we must have

$$\begin{aligned} \hat{\sigma}_1 \, d\hat{W}_t^1 &= 0.1 \, dW_t^1 + 0.4 \, dW_t^2 \\ \hat{\sigma}_2 \, d\hat{W}_t^2 &= 0.4 \, dW_t^1 + 0.3 \, dW_t^2 \end{aligned}$$

Looking at the covariance processes, we must have

$$\begin{aligned} \hat{\sigma}_1^2 \, dt &= (0.1^2 + 0.4^2) \, dt \\ \hat{\sigma}_2^2 \, dt &= (0.4^2 + 0.3^2) \, dt \\ \hat{\sigma}_1 \hat{\sigma}_2 \rho \, dt &= (0.1 \times 0.4 + 0.4 \times 0.3) \, dt \end{aligned}$$

which are three equations in 3 unknowns, easily solved for $\hat{\sigma}_1, \hat{\sigma}_2, \rho$:

$$\begin{aligned} \hat{\sigma}_1 &= ||(0.1, 0.4)|| \\ \hat{\sigma}_2 &= ||(0.4, 0.3)|| \\ \rho &= \frac{(0.1, 0.4) \cdot (0.4, 0.3)}{||(0.1, 0.4)|| \cdot ||(0.4, 0.3)||} \end{aligned}$$

Note that the vectors on the right can all be read off the volatility matrix.

Thus

$$\hat{W}_t^1 = \frac{(0.1, 0.4) \cdot (W_t^1, W_t^2)}{||(0.1, 0.4)||}$$
$$\hat{W}_t^2 = \frac{(0.4, 0.3) \cdot (W_t^1, W_t^2)}{||(0.4, 0.3)||}$$

It is clear that $\hat{W}^1, \hat{W}^2$ are continuous martingales. Moreover

$$[\hat{W}^1]_t = t = [\hat{W}^2]_t$$

so that $\hat{W}^1, \hat{W}^2$ are indeed Brownian motions (by Lévy's Characterization). Furthermore,

$$[\hat{W}^1, \hat{W}^2]_t = \rho \, dt$$

as expected.

□

The above example can be generalized:

Proposition 7.4.3 *Give a multidimensional SDE $dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$, i.e.*

$$\begin{pmatrix} dX_t^1 \\ \vdots \\ dX_t^n \end{pmatrix} = \begin{pmatrix} b^1(t, X_t) \\ \vdots \\ b^n(t, X) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(t, X_t) & \dots & \sigma_{1m}(t, X_t) \\ \vdots & \vdots & \vdots \\ \sigma_{n1}(t, X_t) & \dots & \sigma_{nm}(t, X_t) \end{pmatrix} \begin{pmatrix} dW_t^1 \\ \vdots \\ dW_t^m \end{pmatrix}$$

where $W_t = (W_t^1, \dots, W^m)_t$ is a standard m–dimensional Brownian motion. Let σ_i be the i^{th} row of the matrix σ. Define

$$\hat{W}_t^i = \frac{\sigma_i \cdot W_t}{||\sigma_i||} \qquad \text{for } i = 1, \dots, n$$

Then (by Lévy's Characterization) the $\hat{W}_t^i$ are n correlated Brownian motions, with correlation

$$\rho_{ij} = \frac{\sigma_i \cdot \sigma_j}{||\sigma_i|| \cdot ||\sigma_j||}$$

and we have dynamics

$$dX_t^i = b^i(t, X_t)\, dt + ||\sigma_i(t, X_t)||\, d\hat{W}_t^i \qquad \text{for } i = 1, \dots, n$$

Here each X^i is driven by only one source of noise.

□

Thus the "volatility" of a one–dimensional process of the form

$$dX_t = \mu\, dt + \sigma_1\, dW^1 + \dots + \sigma_n\, dW^n$$

is

$$\hat{\sigma} = ||(\sigma_1, \dots, \sigma_n)|| = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

What happens to the Black–Scholes PDE when we have correlated Brownian motions? Recall that this is

$$\frac{\partial C}{\partial t} + \sum_n rS^n \frac{\partial C}{\partial S^n} + \frac{1}{2}\sum_{n,m}(\sigma\sigma^{tr})S^n S^m \frac{\partial^2 C}{\partial S^n \partial S^m} - rC = 0$$

Note that $(\sigma\sigma^{tr})_{nm}$ is just $\sigma_n \cdot \sigma_m$, the inner product of the n^{th} and m^{th} rows of σ. We have seen that

$$\rho_{nm} = \frac{\sigma_n \cdot \sigma_m}{||\sigma_n||||\sigma_m||} = \frac{\sigma_n \cdot \sigma_m}{\hat{\sigma}_n \hat{\sigma}_m}$$

and thus we obtain

$$\frac{\partial C}{\partial t} + \sum_n rS^n \frac{\partial C}{\partial S^n} + \frac{1}{2}\sum_{n,m}\rho_{nm}\hat{\sigma}_n\hat{\sigma}_m S^n S^m \frac{\partial^2 C}{\partial S^n \partial S^m} - rC = 0$$

where $\hat{\sigma}_n$ is the volatility of S^n.

7.5 Change of Numéraire

7.5.1 Introduction to Change of Numéraire

Thus far, we've mainly considered two probability measures, the "real world" measure $\mathbb{P}$, and the equivalent martingale measure $\mathbb{Q}$ for the money market account numéraire. We've seen, however, that it is possible to introduce an EMM for different numéraires, and to use these for pricing. We now show that a change of numéraire is a technique which often simplifies a pricing problem – it is analogous to a reduction in dimension.

Consider an interest rate derivative X, and let A_t be the bank account. If $\mathbb{Q}$ is the EMM for A_t, then $\frac{X_t}{A_t}$ is a $\mathbb{Q}$–martingale (assuming, of course, that X is attainable).Thus

$$X_0 = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r(s,\omega)\,ds}\, X(T)\right]$$

where r is the short rate, so that $A_t = e^{\int_0^t r(s,\omega)\,ds}$. In order to compute this, we would have to know the joint density of $X(T)$, $A(T)$ under $\mathbb{Q}$ — it would not be observable, because only $\mathbb{P}$–densities can be observed. The computation of the expectation would involve a double integral.

The reason you may not have noticed this problem before is that we have generally assumed that interest rates are constant, which simplifies matters considerably. If we assume that the payoff $X(T)$ and the short rate are independent under $\mathbb{Q}$, then we would still have some simplification, namely

$$\begin{aligned} X_0 &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r(s,\omega)\,ds}\right]\mathbb{E}_{\mathbb{Q}}\left[X(T)\right] \\ &= p(0,T)\mathbb{E}_{\mathbb{Q}}\left[X(T)\right] \end{aligned}$$

where $p(t,T)$ is the time t–price of a zero coupon bond with face value 1 and maturity T, i.e. an interest rate derivative with payoff 1 at expiry, in all states of the world. The above expression is obviously much simpler:

- It only involves a single integral, and needs only the $\mathbb{Q}$–density of $X(T)$.

- $p(0, T)$ is observable (either directly, or by bootstrapping a yield curve from observable coupon bond prices).

Generally, of course, $X(T), A(T)$ are *not* independent under $\mathbb{Q}$. Even if they *were* independent under $\mathbb{P}$, they would nevertheless probably not be independent under $\mathbb{Q}$ — under $\mathbb{Q}$, the drifts of all assets are the same, namely the short rate. Thus X_t has the same drift as A_t, implying some correlation.

7.5.2 Mechanics of Changes of Numéraire

As usual, we work with a market model $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t, (S_t^0, \ldots, S_t^N)_t)$. Recall:

- A *numéraire* is a traded asset (posibly a portfolio of assets) with a strictly positive price process.
- Self–financing portfolios remain self–financing under a change of numéraire.
- Replicating portfolios remain replicating portfolios.

Now the bank account $S_t^0 = A_t$ is just a special numéraire — one whose dynamics have zero volatility: $dA_t = r(t, \omega) A_t \, dt$. Let $\mathbb{Q}$ be the EMM for A_t. Then each $\frac{S_t^n}{A_t}$ is a $\mathbb{Q}$–martingale, i.e. $\mathbb{Q}$ "martingalizes" the ratios $\frac{S_t^n}{A_t}$.

Suppose that $\hat{A}(t)$ is another numéraire, with EMM $\hat{\mathbb{Q}}$. $\hat{\mathbb{Q}}$ "martingalizes" the ratios $\frac{S_t^n}{\hat{A}_t}$. Given that $\hat{A}(t)$ is a (combination of) traded assets, we expect $\frac{\hat{A}(t)}{A(t)}$ to be a $\mathbb{Q}$–martingale as well. If X is an attainable claim, then $\frac{X_t}{A_t}$ is a $\mathbb{Q}$–martingale, and $\frac{X_t}{\hat{A}_t}$ is a $\hat{\mathbb{Q}}$–martingale.

What does $\hat{\mathbb{Q}}$ look like? Since $\mathbb{Q}, \hat{\mathbb{Q}}$ are both equivalent to $\mathbb{P}$, they are equivalent to each other, and thus the Radon–Nikodym derivative

$$L(T) = \frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}$$

exists. We don't know yet what it is, though, because we don't know $\hat{\mathbb{Q}}$. Nevertheless, it exists, so we may define the likelihood process

$$L(t) = \mathbb{E}_{\mathbb{Q}}[L(T)|\mathcal{F}_t]$$

We have, by Bayes' Theorem,

$$\frac{X_0}{\hat{A}_0} = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{X(T)}{\hat{A}(T)}\right] = L(0)^{-1}\mathbb{E}_{\mathbb{Q}}\left[\frac{X(T)}{A(T)}L(T)\right]$$

so that

$$\frac{X_0}{A_0} = \mathbb{E}_{\mathbb{Q}}\left[\frac{X(T)}{A(T)}\frac{\hat{A}_0}{A_0}L(T)\right]$$

because $L(0) = 1$. But

$$\frac{X_0}{A_0} = \mathbb{E}_{\mathbb{Q}}\left[\frac{X(T)}{A(T)}\right]$$

This suggests that we turn every thing around and *define*

$$L(T) = \frac{\hat{A}(T)/A(T)}{\hat{A}(0)/A(0)}$$

and then *define* $\hat{\mathbb{Q}}$ by

$$\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}} = L(T)$$

Then $L(t) = \frac{\hat{A}_t/A_t}{\hat{A}_0/A_0}$, as you can easily check.

In general, we may use for $\hat{A}(t)$ absolutely any process with the property that $\frac{\hat{A}_t}{A_t}$ is a strictly positive $\mathbb{Q}$–martingale.

Theorem 7.5.1 (Martingale Measure Pricing)
Suppose that $\hat{A}(t)$ is process with the property that $\frac{\hat{A}_t}{A_t}$ is a strictly positive $\mathbb{Q}$–martingale. Define

$$L(t) = \frac{\hat{A}_t/A_t}{\hat{A}_0/A_0} \qquad \frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}} = L(T)$$

If $\frac{M_t}{A_t}$ is a $\mathbb{Q}$–martingale, then $\frac{M_t}{\hat{A}_t}$ is a $\hat{\mathbb{Q}}$–martingale. In particular, if X is an attainable contingent claim, then

$$X_t = \hat{A}_t \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{X(T)}{\hat{A}(T)}|\mathcal{F}_t\right]$$

□

In fact, we can generalize even more:

Theorem 7.5.2 *Suppose that $\alpha_1(t), \alpha_2(t)$ are numeraires, and that $\mathbb{Q}_1, \mathbb{Q}_2$ are their associated EMM's. Then for any random variable X we have*

$$\alpha_1(t)\mathbb{E}_{\mathbb{Q}_1}\left[\frac{X}{\alpha_1(T)}|\mathcal{F}_t\right] = \alpha_2(t)\mathbb{E}_{\mathbb{Q}_2}\left[\frac{X}{\alpha_2(T)}|\mathcal{F}_t\right]$$

Proof: Define the likelihood process $L_1(t) = \frac{\alpha_1(t)/A(t)}{\alpha_1(0)/A(0)}$, and define $L_2(t)$ similarly. Then by Bayes' Theorem

$$\begin{aligned}\alpha_1(t)\mathbb{E}_{\mathbb{Q}_1}\left[\frac{X}{\alpha_1(T)}|\mathcal{F}_t\right] &= \alpha_1(t)L_1(t)^{-1}\mathbb{E}_{\mathbb{Q}}\left[\frac{X}{\alpha_1(T)}L_1(T)|\mathcal{F}_t\right] \\ &= A(t)\mathbb{E}_{\mathbb{Q}}\left[\frac{X}{A(T)}|\mathcal{F}_t\right]\end{aligned}$$

and the same goes for α_2.

□

Let's investigate how asset price dynamics change when we move from $\mathbb{Q}$–world to $\hat{\mathbb{Q}}$–world. Assuming that asset prices are Itô diffusions, we have $\mathbb{Q}$–dynamics

$$dS_t = D[S_t]r_t\ dt + D[S_t]\sigma_t\ dW_t$$

where W is a (K–dimensional) $\mathbb{Q}$–Brownian motion. The Radon–Nikodym derivative (process) which effects the change from $\mathbb{Q}$ to $\hat{\mathbb{Q}}$ is

$$L(t) = \frac{\hat{A}(t)/A(t)}{\hat{A}(0)/A(0)}$$

Using the fact that $\hat{A}_t/A_t$ is a $\mathbb{Q}$–martingale, we see that

$$d\hat{A}_t = r_t\hat{A}_t\ dt + \hat{\sigma}_t\hat{A}_t\ dW_t$$

By Itô's formula,

$$dL_t = \frac{A_0}{\hat{A}_0}\left[\frac{1}{A_t}\left(r_t\hat{A}_t\ dt + \hat{\sigma}_t\hat{A}_t\ dW_t\right) - \frac{\hat{A}_t}{A_t^2}\left[r_tA_t\ dt\right)\right]$$

Thus

$$dL_t = L_t\hat{\sigma}_t\ dW_t$$

confirming that $L(t)$ is a $\mathbb{Q}$–martingale, as we already knew. Solving the SDE, we obtain

$$L(t) = e^{\int_0^t \hat{\sigma}_s\ dW_s - \frac{1}{2}\int_0^t ||\hat{\sigma}_s||^2\ ds}$$

and thus

> Suppose we change the numéraire from the MMA A_t to $\hat{A}_t$. Then the EMM $\hat{\mathbb{Q}}$ associated with $\hat{A}_t$ is obtained from the EMM $\mathbb{Q}$ associated with A_t by a Girsanov transformation whose kernel is the volatility $\hat{\sigma}$ of the new numéraire $\hat{A}_t$.

By Girsanov's Theorem

$$\hat{W}_t = W_t - \int_0^t \hat{\sigma}_s\ ds$$

is a (K–dimensional) $\hat{\mathbb{Q}}$–Brownian motion, and thus the asset dynamics under $\hat{\mathbb{Q}}$ are given by

$$dS_t = D[S_t]\begin{pmatrix} r_t + \sigma_t^1 \cdot \hat{\sigma}_t \\ \vdots \\ r_t + \sigma_t^N \cdot \hat{\sigma}_t \end{pmatrix} dt + D[S_t]\sigma\ d\hat{W}_t$$

i.e.

$$dS_t^n = (r_t + \sigma_t^n \cdot \hat{\sigma})S_t^n\ dt + \sigma_t^n S_t^n\ d\hat{W}_t$$

where σ^n is the n^{th} row of the volatility matrix σ. In particular, the "discounted" asset ratios $\hat{S}_t^n = \frac{S_t^n}{\hat{A}_t}$ have dynamics

$$d\hat{S}_t^n = \hat{S}_t^n(\sigma_t^n - \hat{\sigma}_t)\ d\hat{W}^t$$

as you can easily verify by applying Itô's formula. Hence the $\hat{S}_t^n$ are $\hat{\mathbb{Q}}$–martingales, and the volatility of each "discounted" asset is reduced by the volatility of the numéraire.

Remarks 7.5.3 Consider a simple Black–Scholes model, where the risky asset prices are given by a geometric Brownian motions, driven by a single source of noise. The market price of risk of S^n in $\hat{\mathbb{Q}}$–world is

$$\frac{r + \sigma^n \hat{\sigma} - r}{\sigma^n} = \hat{\sigma}$$

i.e. all assets have the same market price of risk, namely the volatility of the numéraire. This is also true in the multidimensional case, where the market price of risk is a vector.

The bank account has zero volatility, and thus the MPR in $\mathbb{Q}$–world is zero.

□

7.5.3 A General Option pricing Formula

Consider a call C on a security S with strike K and maturity T. Let $\mathbb{Q}_S$ be the EMM associated with numéraire S, and let $\mathbb{Q}^T$ be the T–forward measure (i.e. the EMM associated with the zero coupon bond $p(t,T)$ maturing at T).

Theorem 7.5.4

$$C_0 = S_0 \mathbb{Q}_S(S_T \geq K) - KP(0,T)\mathbb{Q}^T(S_T \geq K)$$

Proof: We have

$$\begin{aligned} C_0 &= p(0,T)\mathbb{E}_{\mathbb{Q}^T}[\max\{S_T - K, 0\}] \\ &= p(0,T)\mathbb{E}_{\mathbb{Q}^T}[S_T - K; S_T \geq K] \\ &= p(0,T)\mathbb{E}_{\mathbb{Q}^T}[S_T; S_T \geq K] - Kp(0,T)\mathbb{Q}^T(S_T \geq K) \end{aligned}$$

But we have

$$\alpha_1(t)\mathbb{E}_{\mathbb{Q}_1}\left[\frac{X}{\alpha_1(T)}|\mathcal{F}_t\right] = \alpha_2(t)\mathbb{E}_{\mathbb{Q}_2}\left[\frac{X}{\alpha_2(T)}|\mathcal{F}_t\right]$$

for general numéraires and their associated EMM's. Using this with $\alpha_1(t) = p(t,T)$ and $\alpha_2(t) = S_t$, we obtain

$$\begin{aligned} p(0,T)\mathbb{E}_{\mathbb{Q}^T}[S_T; S_T \geq K] &= p(0,T)\mathbb{E}_{\mathbb{Q}^T}\left[\frac{S_T I_{\{S_T \geq K\}}}{p(T,T)}\right] \\ &= S_0 \mathbb{E}_{\mathbb{Q}_S}\left[\frac{S_T I_{\{S_T \geq K\}}}{S_T}\right] \\ &= S_0 \mathbb{Q}_S(S_T \geq K) \end{aligned}$$

□

7.5.4 Applications

Example 7.5.5 Forward Measures
Consider again the situation at the beginning of this chapter: We consider a contingent claim X with expiry T. Under riskneutral valuation, its value is

$$X_0 = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r(s,\omega)\,ds} X(T)\right]$$

where r is the short rate. We bemoaned the fact that this would necesssitate us knowing the joint density of A_T and X_T. If only, we said, A_T and X_T were independent, we would get the much simpler

$$\begin{aligned} X_0 &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r(s,\omega)\,ds}\right]\mathbb{E}_{\mathbb{Q}}\left[X(T)\right] \\ &= p(0,T)\mathbb{E}_{\mathbb{Q}}\left[X(T)\right] \end{aligned}$$

where $p(t,T)$ is the time t–price of a zero coupon bond with face value 1 and maturity T. If only. . .

Now let's see what happens if we change the numéraire to $p(t,T)$, and let $\mathbb{Q}_T$ be the corresponding EMM. In that case, the pricing formula becomes

$$\frac{X_0}{p(0,T)} = \mathbb{E}_{\mathbb{Q}_T}\left[\frac{X_T}{p(T,T)}\right]$$

and noting that $P(T,T) = 1$, we have

$$X_0 = p(0,T)\mathbb{E}_{\mathbb{Q}_T}\left[X_T\right]$$

This is the simple form that we sought, but it's correct under $\mathbb{Q}_T$, and *not* under $\mathbb{Q}$.

The measure $\mathbb{Q}_T$ is called the *T–forward measure.* Note that if interest rates are deterministic, then $\mathbb{Q}$ and $\mathbb{Q}_T$ coincide, because then $p(t,T) = e^{-\int_t^T r_s\,ds}$, so that each ratio $S_t^n/p(t,T) = A_T(S_t^n/A_t) = \text{const.} \times S_t^n/A_t$ is already a $\mathbb{Q}$–martingale.

However, when interest rates are stochastic, $\mathbb{Q}$ and $\mathbb{Q}_T$ are quite different. We shall see later that futures prices are $\mathbb{Q}$–martingales, whereas forward prices are $\mathbb{Q}_T$–martingales. Thus forward prices and futures prices coincide if interest rates are deterministic.

□

Example 7.5.6 Exchange Options

Consider an *exchange option* which gives the right, but not the obligation, to exchange asset S^1 for asset S^2 at time T. This is a contingent claim X with payoff

$$X_T = \max\{S_T^2 - S_T^1, 0\}$$

Using riskneutral valuation (i.e. MMA as numéraire), its value is therefore

$$X_0 = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}\max\{S_T^2 - S_T^1, 0\}\right]$$

To compute this, we have to know the joint distributions of A_T, S_T^1, S_T^2 under $\mathbb{Q}$, yielding a triple integral.

It is computationally simpler to change the numéraire: Let $\hat{A}_t = S_T^1$, and let $\hat{\mathbb{Q}}$ be the associated EMM. The contingent claim is then priced as follows:

$$\frac{X_0}{S_0^1} = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{\max\{S_T^2 - S_T^1, 0\}}{S_T^1}\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\max\{\hat{S}_T^2 - 1, 0\}\right]$$

(where $\hat{S}_t^2 = \frac{S_t^2}{S_t^1}$). This looks like a call on $\hat{S}^2$ with strike $K = 1$, and we only have to know the distribution of $\hat{S}_T^2$ under $\hat{\mathbb{Q}}$.

To price this option, we have to assume some form of asset dynamics. Suppose these are given by one–dimensional Itô diffusions, i.e. suppose we have $\mathbb{P}$–dynamics

$$\begin{aligned} dS_t^1 &= \mu_1 S_t^1 \, dt + \bar{\sigma}_1 \, d\bar{W}_{\mathbb{P}}^1(t) \\ dS_t^2 &= \mu_2 S_t^2 \, dt + \bar{\sigma}_2 \, d\bar{W}_{\mathbb{P}}^2(t) \end{aligned}$$

where $\bar{W}_{\mathbb{P}}^1, \bar{W}_{\mathbb{P}}^2$ are *correlated* $\mathbb{P}$–Brownian motions, with correlation ρ_t. To get lognormality we also assume that $\bar{\sigma}_1(t), \bar{\sigma}_2(t)$ and $\rho(t)$ are *deterministic*. Interest rates, however, can be stochastic.

We can write this as

$$d\begin{pmatrix} S_t^1 \\ S_t^2 \end{pmatrix} = \begin{pmatrix} \mu_1 S_t^1 \\ \mu_2 S_t^2 \end{pmatrix} dt + D[S_t]\sigma_t \, dW_{\mathbb{P}}(t)$$

where $W_{\mathbb{P}}^1, W_{\mathbb{P}}^2$ are independent $\mathbb{P}$–Brownian motions and

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Then we must have

$$\begin{aligned} \bar{\sigma}_1 &= \sqrt{\sigma_{11}^2 + \sigma_{12}^2} \\ \bar{\sigma}_2 &= \sqrt{\sigma_{21}^2 + \sigma_{22}^2} \quad \text{and} \\ \rho &= \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\sqrt{\sigma_{11}^2 + \sigma_{12}^2}\sqrt{\sigma_{21}^2 + \sigma_{22}^2}} \end{aligned}$$

So given $\bar{\sigma}_1, \bar{\sigma}_2$ andρ we can solve for the matrix σ (though not uniquely).

Note that the correlation is a function of the volatility matrix. When we change from $\mathbb{P}$–world to $\mathbb{Q}$–world, the volatility matrix is unchanged, and thus also the correlation. Thus under $\mathbb{Q}$, the asset dynamics are

$$\begin{aligned} dS_t^1 &= r_t S_t^1 \, dt + \bar{\sigma}_1 \, d\bar{W}_{\mathbb{Q}}^1(t) \\ dS_t^2 &= r_t S_t^2 \, dt + \bar{\sigma}_2 \, d\bar{W}_{\mathbb{Q}}^2(t) \end{aligned}$$

where $\bar{W}_{\mathbb{Q}}^1, \bar{W}_{\mathbb{Q}}^2$ are $\mathbb{Q}$–Brownian motions, with correlation ρ. This can also be written as

$$d\begin{pmatrix} S_t^1 \\ S_t^2 \end{pmatrix} = D[S_t] r_t \, dt + D[S_t]\sigma_t \, dW_{\mathbb{Q}}(t)$$

where $W_{\mathbb{Q}}^1, W_{\mathbb{Q}}^2$ are independent $\mathbb{Q}$–Brownian motions. Now when we change the numéraire from the MMA to S^1, and the measure from $\mathbb{Q}$ to $\hat{\mathbb{Q}}$, we get

$$\begin{aligned} d\hat{S}_t^2 &= \hat{S}_t^2(\sigma_{21} - \sigma_{11}, \sigma_{22} - \sigma_{12}) \cdot \, d\begin{pmatrix} W_{\hat{\mathbb{Q}}}^1(t) \\ W_{\hat{\mathbb{Q}}}^2(t) \end{pmatrix} \\ &= \hat{S}_t^2 \hat{\sigma}_2 \, d\bar{W}_{\hat{\mathbb{Q}}}(t) \end{aligned}$$

where $\bar{W}_{\hat{\mathbb{Q}}}$ is a one–dimensional $\hat{\mathbb{Q}}$–Brownian motion and

$$\hat{\sigma}_2 = \sqrt{(\sigma_{21} - \sigma_{11})^2 + (\sigma_{22} - \sigma_{12})^2} = \sqrt{(\bar{\sigma}_1)^2 + (\bar{\sigma}_2)^2 - 2\rho\bar{\sigma}_1\sigma_2}$$

Now since the $\bar{\sigma}_1(t), \bar{\sigma}_2(t)$ and $\rho(t)$ are assumed to be deterministic, so is $\hat{\sigma}_2(t)$. It follows that $\hat{S}^2_T$ is lognormal under $\hat{\mathbb{Q}}$: Indeed

$$\hat{S}^2_T = \hat{S}^2_0 e^{\int_0^T \hat{\sigma}_2(t)\, d\bar{W}_{\hat{\mathbb{Q}}}(t) - \frac{1}{2}\int_0^T \hat{\sigma}_2(t)^2\, dt}$$

so that

$$\ln(S^2_T/S^2_0) \sim N\left(-\frac{1}{2}\int_0^T \hat{\sigma}_2(t)^2\, dt, \frac{1}{2}\int_0^T \hat{\sigma}_2(t)^2\, dt\right)$$

Using the properties of lognormality, we see that

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\max\{\hat{S}^2_T - 1, 0\}\right] = \mathbb{E}_{\hat{\mathbb{Q}}}[\hat{S}^2_T]N(d_1) - 1N(d_2)$$

where

$$d_1 = \frac{\ln(\frac{\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{S}^2_T]}{1}) + \frac{1}{2}\int_0^T \hat{\sigma}_2(t)^2\, dt}{\sqrt{\int_0^T \hat{\sigma}_2(t)^2\, dt}}$$
$$d_2 = d_1 - \sqrt{\int_0^T \hat{\sigma}_2(t)^2\, dt}$$

and thus, using the fact that $\mathbb{E}_{\hat{\mathbb{Q}}}[\hat{S}^2_T] = \hat{S}^2_0$, we have

$$X_0 = S^2_0 N(d_1) - S^1_0 N(d_2)$$

If we further assume that $\bar{\sigma}_1, \bar{\sigma}_2$ and ρ are constants, we obtain

$$X_0 = S^2_0 N(d_1) - S^1_0 N(d_2)$$
$$\text{where} \quad d_1 = \frac{\ln(\frac{S^2_0}{S^1_0}) + \frac{1}{2}(\bar{\sigma}_1^2 + \bar{\sigma}_2^2 - 2\rho\bar{\sigma}_1\bar{\sigma}_2)T}{\sqrt{\bar{\sigma}_1^2 + \bar{\sigma}_2^2 - 2\rho\bar{\sigma}_1\bar{\sigma}_2)T}}$$
$$d_2 = d_1 - \sqrt{\bar{\sigma}_1^2 + \bar{\sigma}_2^2 - 2\rho\bar{\sigma}_1\bar{\sigma}_2)T}$$

where we used the fact that

$$\hat{\sigma}_2 = \sqrt{(\bar{\sigma}_1)^2 + (\bar{\sigma}_2)^2 - 2\rho\bar{\sigma}_1\sigma_2}$$

□

Note that, in the above example, $\frac{1}{T}\int_0^T \hat{\sigma}_2(t)^2\, dt$ is just the average of the squared volatility of $\hat{S}^2$, so that $\int_0^T \hat{\sigma}_2(t)^2\, dt = \sigma^2_{\text{average}}T$.

Chapter 8

Interest Rate Modelling

8.1 Modelling Fixed Income: Introduction

8.1.1 Classification of Interest Rate Models

We will examine several approaches for the modelling of interest rates: short rate modelling, whole yield curve modelling and market models. The purpose of this section is to introduce basic concepts and notation. Amongst the immediately obvious quantities that we may model are

- bond prices
- the short rate
- forward rates (discretely or continuously compounded)
- the entire yield curve

Of course, a model of bond prices will have the yield curve as an output, etc. These approaches are no independent.

Short rate models: These model just one variable, the short rate, which is an idealized quantity that represents the instantaneous interest rate at any time. Usually a diffusion model, and thus Markov. We specify dynamics, e.g.

$$dr_t = \kappa(\theta - r_t)\,dt + \sigma\,dW_t \qquad \kappa, \theta, \sigma \text{ const.}$$

is the Vasiček model, and

$$dr_t = \kappa(\theta(t) - r_t)\,dt + \sigma(t)\,dW_t \qquad \kappa \text{ const.}, \theta, \sigma \text{ deterministic}$$

is the Hull–White (extended Vasiček) model.

- Can be one–factor or multi–factor
- *Affine* term structure models have a particularly simple form, allowing for closed form solutions for bond option prices, Eurodollar futures, etc. More later…

- Multi–factor models: principal component analysis shows that 80-90% of the variance of the term structure is explained by parallel shifts of the yield curve, 5-10% by a twist (long term and short term rates move in opposite directions, pivoting about a point), and 1-2% by a butterfly (long and short term rates move in the same direction, with mid–term rates moving in the opposite direction).

Whole yield curve models: These model the entire term structure of rates, eg. the entire forward rate curve. Examples are

- Heath–Jarrow–Morton models
- Market models

Interest rate models are often categorized into *Equilibrium models* and *No–arbitrage models*. Equilibrium models attempt to derive, e.g., short rate dynamics from macroeconomic considerations, starting from a representative investor (e.g. Cox–Ingersoll–Ross, Vasiček, Merton models). These models often have the nice property of being time–homogeneous, but usually are unable to fit observed prices exactly. No–arbitrage models attempt to fit a model exactly to observed prices and volatilities –zero coupon bonds, caplets, swaptions. (e.g. Ho–Lee, Hull–White models).

Both terms are misnomers: Some equilibrium models are not arbitrage–free, and thus *not* in equilibrium. Some no–arbitrage models permit negative interest rates, thus allowing "mattress arbitrage" (borrow from the bank when rates go negative, put under mattress).

8.1.2 Bond Market Basics

One of the basic instruments that we shall be concerned with is the following:

Definition 8.1.1 A T–bond is a zero coupon bond with face value 1.00 and maturity T. Its value at time $t \leq T$ is denoted by $p(t,T)$.

$\square$

These are also called discount bonds.

We work in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration which satisfies the usual conditions. We usually require that:

- $p(t,T)$ is a continuous semimartingale for each T.
- $0 \leq p(t,T) \leq 1$ a.s. (This fails in, e.g. Gaussian short rate models)
- There is a frictionless market for T–bonds of every maturity $T > 0$.
- For every fixed $T > 0$, $\{p(t,T) : 0 \leq t \leq T\}$ is an optional process with $P(T,T) = 1$.
- For every fixed t, $p(t,T)$ is ($\mathbb{P}$–a.s.) differentiable in the second variable T.

$$p_T(t,T) = \frac{\partial p(t,T)}{\partial T}$$

- No default risk.

Note that, for fixed t, the set $\{p(t,T) : T \geq 0\}$ is just the term structure of zero coupon bond prices, which is typically a smooth decreasing function (of T). On the other hand, for fixed T, the set $\{p(t,T) : t \leq T\}$ is the price process of the security $p(t,T)$, which is typically very ragged (i.e. of unbounded variation).

Note that there are, in our model, infinitely many securities, namely one $p(t,T)$ for each maturity T.

We briefly recall the definitions of the various types of rate:

- Let $t < S < T$. Consider the following strategy:
  - (i) At time t, short an S–bond, and use the proceeds to buy $\frac{p(t,S)}{p(t,T)}$–many T–bonds. Net cashflow at time t is zero.
  - (ii) At time S, pay \$1.00 to redeem the S–bond.
  - (iii) At time T, receive $\frac{p(t,S)}{p(t,T)}$ from maturing T–bonds.

  Thus at time T, we can, with no initial cash outlay, ensure that a deposit of \$1.00 at time S leads to a payoff of $\frac{p(t,S)}{p(t,T)}$ at time T. This implies that we can lock in an interest rate $R(t;S,T)$ for the future period $[S,T]$:

$$\begin{aligned} e^{R(T-S)} &= \frac{p(t,S)}{p(t,T)} \\ \Rightarrow R(t;S,T) &= -\frac{\ln p(t,T) - \ln p(t,S)}{T-S} \end{aligned}$$

  This is the *forward rate* (continuously compounded) for the period $[S,T]$ at time t.

- The equivalent *simple forward rate* (the LIBOR forward rate) for $[S,T]$ contracted at time t is similarly defined by

$$\begin{aligned} 1 + L(T-S) &= \frac{p(t,S)}{p(t,T)} \\ \Rightarrow L(t;S,T) &= -\frac{p(t,T) - p(t,S)}{p(t,T)(T-S)} \end{aligned}$$

- The continuously and simple *spot rates* at time t for time T are $R(t;t,T)$ and $L(t;t,T)$ respectively.

- The *instantaneous forward rate* at time t for time T is the interest rate that can be locked in for an infinitesimal interval $[t, T+dT]$. It is given by

$$\begin{aligned} f(t,T) &= \lim_{\Delta t \to 0} R(t;T,T+\Delta T) \\ &= -\lim_{\Delta t \to 0} \frac{\ln p(t,T+\Delta T) - \ln p(t,T)}{\Delta T} \\ &= -\frac{\partial \ln p(t,T)}{\partial T} \end{aligned}$$

- The *short rate* is the instantaneous spot rate, and is defined by

$$r(t) = f(t,t)$$

- Given a *tenor structure*

$$0 \leq t \leq T_0 < T_1 < T_2 < \cdots < T_N$$

we can find a *forward swap rate* $S_t = S(t; T_0, T_1, \ldots, T_N)$, the unique fixed rate, at time t, for which a fixed–for–floating forward swap, starting at time T_0, will have zero value. We clearly require, with $\tau_j = T_j - T_{j-1}$, that

$$\sum_{j=1}^{N} S_t \tau_j p(t, T_j) = \sum_{j=1}^{N} L(t; T_{j-1}, T_j) \tau_j p(t, T_j)$$

and thus

$$S_t = \frac{\sum_{j=1}^{N} L(t; T_{j-1}, T_j) \tau_j p(t, T_j)}{\sum_{j=1}^{N} \tau_j p(t, T_j)}$$

But

$$\sum_{j=1}^{N} L(t; T_{j-1}, T_j) \tau_j p(t, T_j) = \sum_{j=1}^{N} -[p(t, T_j) - p(t, T_{j-1})] = p(t, T_0) - p(t, T_N)$$

and hence

$$S_t = \frac{p(t, T_0) - p(t, T_N)}{\sum_{j=1}^{N} \tau_j p(t, T_j)}$$

The denominator $\sum_{j=1}^{N} \tau_j p(t, T_j)$ is sometimes referred to as the *value of a basis point*.

Remarks 8.1.2 1. The assumption that there are traded zero coupon bonds of every maturity is clearly false. Nevertheless, a large number of *implied* zero coupon bond prices can usually be obtained by bootstrapping the yield curve.

2. The instantaneous rates (forward– and short–) are *theoretical* entities, and not directly observable in the market. One of the shortcomings of short rate and HJM models is that they model these non-existent entities. *Market models* such as the BGM– and Jamshidian models, however, are concerned with the modelling of quoted market rates.

□

The following lemma shows how bond prices are related to forward rates:

Lemma 8.1.3

$$p(t, T) = p(t, S) e^{-\int_S^T f(t,u)\, du}$$

Proof: $\ln p(t, T) = \ln p(t, S) + \int_S^T \frac{\partial \ln p(t,u)}{\partial T}\, du$.

□

As usual, we denote that money market account (MMA) process A_t by

$$A_t = e^{\int_0^t r(u)\, du}$$

where $r(t)$ is the short rate.

Example 8.1.4 *No model that allows only parallel shifts of the yield curve is arbitrage–free.*
Proof: Suppose it is certain that $f(1,T) = f(0,T) + \varepsilon$ for all $T \geq 1$, where ε is a random variable. Now choose times $1 < T_1 < T_2 < T_3$. At $t = 1$,

$$p(1,T) = e^{-\int_1^T f(1,u)\,du} = e^{-\int_1^T f(0,u)+\varepsilon\,du} = \frac{p(0,T)}{p(0,1)} e^{-\varepsilon(T-1)}$$

Now suppose that we hold x_i T_i–bonds ($i = 1,2,3$). We construct an arbitrage, a static portfolio satisfying

(i) $\sum_{i=1}^3 x_i p(0,T_i) = 0$

(ii) $\sum_{i=1}^3 x_i p(1,T_i) > 0$ a.s.

At time 1 the value of the portfolio is

$$\begin{aligned} V_1(\varepsilon) &= \sum_{i=1}^3 x_i p(1,T_i) \\ &= \sum_{i=1}^3 x_i \frac{p(0,T_i)}{p(0,1)} e^{-\varepsilon(T_i-1)} \\ &= \sum_{i=1}^3 x_i \frac{p(0,T_i)}{p(0,1)} e^{-\varepsilon(T_i-T_2)} e^{-\varepsilon(T_2-1)} \\ &= g(\varepsilon)\frac{e^{-\varepsilon(T_2-1)}}{p(0,1)} \end{aligned}$$

where

$$g(\varepsilon) = \sum_{i=1}^3 x_i p(0,T_i) e^{-\varepsilon(T_i-T_2)}$$

We shall ensure that $V_1(\varepsilon) > 0$ whenever $\varepsilon \neq 0$. First note that $g(0) = 0$, because $\sum_{i=1}^3 x_i p(0,T_i) = 0$. Further, $V_1(\varepsilon)$ and $g(\varepsilon)$ always have the same sign, so to ensure $V_1(\varepsilon) > 0$, it suffices to ensure that $g(\varepsilon) > 0$.

Now g is a C^2–function (twice differentiable), and we require that (i) $g(0) = 0$, (ii) $g(\varepsilon) > 0$ whenever $\varepsilon \neq 0$. It follows that $g'(0) = 0$, thus that

$$g'(0) = \sum_{i=1}^3 x_i (T_2 - T_i) p(0,T_i) = 0$$

and thus that

$$\sum_{i=1}^3 x_i T_i p(0,T_i) = 0$$

Next, if we ensure $g''(\varepsilon) > 0$, then, combined with $g(0) = g'(0) = 0$, we se that $g(\varepsilon) > 0$ for all $\varepsilon \neq 0$. Now

$$g''(\varepsilon) = \sum_{i=1}^3 x_i (T_2 - T_i)^2 p(0,T_i) e^{-\varepsilon(T_i-T_2)}$$

and thus $g''(\varepsilon) > 0$ for all ε if $x_1, x_3 \geq 0$ (and at least one is > 0).

Now take $x_2 < 0$. Since $\sum_{i=1}^{3} x_i p(0, T_i) = 0$, we see that at least one of x_1, x_3 must be > 0. Since $\sum_{i=1}^{3} x_i(T_2 - T_i)p(0, T_i)p(0, T_i) = 0$, we see that x_1, x_3 have the same sign, i.e. both are > 0. Then $g''(\varepsilon) > 0$ for all $\varepsilon \neq 0$, and hence also $g(\varepsilon) > 0$.

It follows that any portfolio (x_1, x_2, x_3) satisfying

(i) $\sum_{i=1}^{3} x_i p(0, T_i) = 0$

(ii) $\sum_{i=1}^{3} x_i(T_2 - T_i)p(0, T_i) = 0$

(iii) $x_2 < 0$

is an arbitrage.

□

Example 8.1.5 Define the *long rate* $l(t)$ by

$$l(t) = \lim_{T \to \infty} R(t, T)$$

where $R(t, T)$ is the c.c. spot rate, i.e. $p(t, T) = e^{-R(t,T)}(T - t)$. Though $l(t)$ is not directly obtainable form traded securities (because the longest–term securities typically have a life of 30 years or so), it can be estimated, and empirical studies suggest that it fluctuates considerably over time. Most no–arbitrage models have a constant value for $l(t)$, however, and indeed

Theorem: *If the term–structure dynamics are arbitrage–free, then $l(t)$ is an increasing function a.s.*

Proof: By rescaling time, we may assume that $l(1) < l(0)$ with positive probability, to obtain a contradiction. For $T = 1, 2, 3, \ldots$, construct a portfolio which, at $t = 0$ invests $\frac{1}{T} - \frac{1}{T+1} = \frac{1}{T(T+1)}$ into each of the bonds $p(t, T)$, so that the value of then portfolio is $V_0 = \sum_{T=1}^{\infty} \frac{1}{T(T+1)} = 1$. Define $\varepsilon = (l(0) - l(1))/3$. Now $p(0, T) = e^{-r(0,T)T}$, and $r(0, T) \to l(0)$ as $T \to \infty$, so eventually, we have $r(0, T) > l(0, T) - \varepsilon$, i.e. $p(0, T) < e^{-(l(0)-\varepsilon)T}$ eventually. Similarly, $p(1, T) > e^{-(l(1)+\varepsilon)T}$ eventually. Suppose these relations hold for all $T \geq T_0$. Then

$$V_1 = \sum_{T=1}^{\infty} \frac{p(1, T)}{T(T+1)p(0, T)} > \sum_{T=1}^{T_0-1} \frac{p(1, T)}{T(T+1)p(0, T)} + \sum_{T=T_0}^{\infty} \frac{e^{\varepsilon T}}{T(T+1)}$$

The second term diverges to ∞, so that $V_1 = \infty$. Now since $V_0 = \mathbb{E}_{\mathbb{Q}}[V_1/B_1]$, where $\mathbb{Q}$ is a risk–neutral measure and B is the bank account, we see that $\mathbb{Q}(V_1 = \infty) = 0$, because $V_0 = 1 < \infty$. Since the "real–world" measure $\mathbb{P}$ is equivalent to $\mathbb{Q}$, we must have $\mathbb{P}(V_1 = \infty) \leq \mathbb{P}(l(1) > l(0)) = 0$ as well.

□

8.1.3 Modelling the Bond Market

We consider three approaches:

1. Specify short rate dynamics;

2. Specify bond price dynamics;

3. Specify forward rate dynamics;

Suppose, for example, that we are given the following dynamics:

1. *Short rate dynamics:*
$$dr(t,\omega) = a(t,\omega)\ dt + b(t,\omega)\ dW_t$$

2. *Bond price dynamics:*
$$dp(t,T)(\omega) = p(t,T)(\omega)[m(t,T,\omega)\ dt + v(t,T,\omega)\ dW_t]$$

3. *Forward rate dynamics:*
$$df(t,T)(\omega) = \alpha(t,T,\omega)\ dt + \sigma(t,T,\omega)\ dW_t$$

Here W_t is a standard (multidimensional) Brownian motion.

If we're given one type of dynamics, can we deduce the others? If you think about this for a while, you'd expect that bond prices and short rates are deduceable from the forward rates, and that forward rates and the short rate are deduceable from the bond prices. A model of just the short rate seems to contain too little information to deduce all bond prices and forward rates however.

Before we write down exactly how the various dynamics are related to each other, we need a stochastic Fubini Theorem and its corollary.

Proposition 8.1.6 (Fubini's Theorem for Stochastic Integrals)
$$\int_0^t \int_0^T \Phi(s,S,\omega)\ dS\ dW_s(\omega) = \int_0^T \int_0^t \Phi(s,S,\omega)\ dW_s(\omega)\ dS$$
where $(s,\omega,S) \mapsto \Phi(s,S,\omega)$ is $\mathcal{P}\times\mathcal{B}$–measurable ($\mathcal{P}$ = predictable σ–algebra, $\mathcal{B}$ = Borel algebra), and

(i) $\int_0^t \Phi^2(s,S,\omega)\ ds < \infty$ *a.s. for all* $t \in [0,T]$*;*

(ii) $\int_0^t \left(\int_0^T \Phi(s,S,\omega)\ dS\right)^2\ ds < \infty$ *a.s. for all* $t \in [0,T]$*;*

(iii) $t \mapsto \int_0^T \int_0^t \Phi(s,S,\omega)\ dW_s(\omega)\ dS$ *is continuous.*

□

The proof is omitted, but may be found in Durrett, Chapter 2, Section 11.

Before we prove a corollary about the differentiation of stochastic integrals, it is convenient to gather well–known results about the differentiation of ordinary Lebesgue integrals:

Proposition 8.1.7 *Assuming sufficient smoothness and regularity,*
$$\frac{\partial}{\partial x}\int_a^x f(y)\ dy = f(x)$$
$$\frac{\partial}{\partial x}\int_a^b f(x,y)\ dy = \int_a^b \frac{\partial}{\partial x} f(x,y)\ dy$$
$$\frac{\partial}{\partial x}\int_{g(x)}^{h(x)} f(x,y)\ dy = \int_{g(x)}^{h(x)} \frac{\partial}{\partial x} f(x,y)\ dy + f(x,h(x))\frac{dh}{dx} - f(x,g(x))\frac{dg}{dx}$$

□

Corollary 8.1.8 (Differentiation under the integral sign)

$$\frac{\partial}{\partial T}\int_0^t v(s,T)\,dW_s = \int_0^t \frac{\partial v(s,T)}{\partial T}\,dW_s$$

Proof: Just like the ordinary proof of differentiation under the integral sign:

$$\begin{aligned}\frac{\partial}{\partial T}\int_0^t v(s,T)\,dW_s &= \frac{\partial}{\partial T}\left[\int_0^t v(s,0) + \int_0^T \frac{\partial v(s,u)}{\partial T}\,du\,dW_s\right]\\ &= 0 + \frac{\partial}{\partial T}\int_0^t\int_0^T \frac{\partial v(s,u)}{\partial T}\,du\,dW_s\\ &= \frac{\partial}{\partial T}\int_0^T\int_0^t \frac{\partial v(s,u)}{\partial T}\,dW_s\,du\\ &= \int_0^t \frac{\partial v(s,T)}{\partial T}\,dW_s\end{aligned}$$

□

Consider now the various dynamics given above, i.e. short rate, bond price and forward rate dynamics. Assume that the drifts and variance rates are C^1 in the T–variable, and sufficiently regular to allow the interchange of order of integration. Further, assume that bond prices are bounded.

The following theorem records the relationships between the various dynamics:

Theorem 8.1.9 *(a) If*

$$\frac{dp}{p} = m\,dt + v\,dW_t$$

then

$$df = \alpha\,dt + \sigma\,dW_t$$

where

$$\begin{aligned}\alpha(t,T) &= v_T(t,T)\cdot v(t,T) - m_T(t,T)\\ \sigma(t,T) &= -v_T(t,T)\end{aligned}$$

(b) If

$$df = \alpha\,dt + \sigma\,dW_t$$

then

$$dr = a\,dt + b\,dW_t$$

where

$$\begin{aligned}a(t) &= f_T(t,t) + \alpha(t,t)\\ b(t) &= \sigma(t,t)\end{aligned}$$

(c) If

$$df = \alpha\ dt + \sigma\ dW_t$$

then

$$\frac{dp}{p} = \left[r(t) + A(t,T) + \frac{1}{2}||S(t,T)||^2\right]\ dt + S(t,T)\ dW_t$$

where

$$A(t,T) = -\int_t^T \alpha(t,s)\ ds$$
$$S(t,T) = -\int_t^T \sigma(t,s)\ ds$$

Here $||\cdot||$ *is just the usual Euclidean norm.*

Before we begin the proof, note that for each T we have a separate security $p(t,T)$, i.e. for each T we have a separate process $(p(t,T))_{t\geq 0}$. It is to these processes that we apply Itô's formula, etc.

Proof: (1) $d\ln p = [m - \frac{1}{2}v^2]\ dt + v\ dW_t$ and thus

$$\ln p(t,T) = \ln p(0,T) + \int_0^t m(s,T) - \frac{1}{2}v^2(s,T)\ dt + \int_0^t v(s,T)\ dW_s$$

so that

$$-f(t,T) = \frac{\partial \ln p(t,T)}{\partial T} = \frac{\partial \ln p(0,T)}{\partial T} + \int_0^t m_T - v_T \cdot v\ ds + \int_0^t v_T\ dW_s$$

Taking differentials yields the result.
(2)

$$\begin{aligned} r(t) = f(t,t) = f(0,t) + \int_0^t \alpha(s,t)\ ds + \int_0^t \sigma(s,t)\ dW_s \quad \text{where} \\ \alpha(s,t) = \alpha(s,s) + \int_s^t \alpha_T(s,u)\ du \\ \sigma(s,t) = \sigma(s,s) + \int_s^t \sigma_T(s,u)\ du \end{aligned}$$

and thus

$$\begin{aligned} r(t) &= f(0,t) + \int_0^t \alpha(s,s)\ ds + \int_0^t\int_s^t \alpha_T(s,u)\ du\ ds + \int_0^t \sigma(s,s)\ dW_s + \int_0^t\int_s^t \sigma_T(s,u)\ du\ dW_s \\ &= f(0,t) + \int_0^t \alpha(s,s)\ ds + \int_0^t\int_0^u \alpha_T(u,s)\ ds\ du + \int_0^t \sigma(s,s)\ dW_s + \int_0^t\int_0^u \sigma_T(u,s)\ dW_s\ du \end{aligned}$$

by the stochastic Fubini theorem. Thus

$$\begin{aligned} dr(t) &= \left[\alpha(t,t) + \int_0^t \alpha_T(s,t)\ ds + \int_0^t \sigma_T(s,t)\ dW_s\right]\ dt + \sigma(t,t)\ dW_t \\ &= [\alpha(t,t) + f_T(t,t)]\ dt + \sigma(t,t)\ dW_t \end{aligned}$$

as required
(3) First define $Y(t,T) = -\int_t^T f(t,s)\, ds$, so that $p(t,T) = e^{Y(t,T)}$. Now

$$f(t,s) = f(0,s) + \int_0^t \alpha(u,s)\, du + \int_0^t \sigma(u,s)\, dW_s$$

and hence

$$\begin{aligned}
Y(t,T) &= -\int_t^T f(0,s)\, ds - \int_t^T \int_0^t \alpha(u,s)\, du\, ds - \int_t^T \int_0^t \sigma(u,s)\, dW_u\, ds \\
&= -\int_t^T f(0,s)\, ds - \int_0^t \int_t^T \alpha(u,s)\, ds\, du - \int_0^t \int_t^T \sigma(u,s)\, ds\, dW_u \\
&= \left[-\int_0^T f(0,s)\, ds + \int_0^t f(0,s)\, ds\right] \\
&\qquad + \left[-\int_0^t \int_u^T \alpha(u,s)\, ds\, du + \int_0^t \int_u^t \alpha(u,s)\, ds\, du\right] \\
&\qquad + \left[-\int_0^t \int_u^T \sigma(u,s)\, ds\, dW_u + \int_0^t \int_u^t \sigma(u,s)\, ds\, dW_u\right] \\
&= Y(0,T) - \int_0^t \int_u^T \alpha(u,s)\, ds\, du - \int_0^t \int_u^T \sigma(u,s)\, ds\, dW_u \\
&\qquad + \left[\int_0^t f(0,s)\, ds + \int_0^t \int_u^t \alpha(u,s)\, ds\, du + \int_0^t \int_u^t \sigma(u,s)\, ds\, dW_u\right] \\
&= Y(0,T) - \int_0^t \int_u^T \alpha(u,s)\, ds\, du - \int_0^t \int_u^T \sigma(u,s)\, ds\, dW_u + \int_0^t f(s,s)\, ds
\end{aligned}$$

Hence

$$Y(t,T) = Y(0,T) + \int_0^t r(s)\, ds - \int_0^t \int_u^T \alpha(u,s)\, ds\, du - \int_0^t \int_u^T \sigma(u,s)\, ds\, dW_u$$

so that

$$\begin{aligned}
dY(t,T) &= \left[r(t) - \int_t^T \alpha(t,s)\, ds\right] dt - \left[\int_t^T \sigma(t,s)\, ds\right] dW_t \\
&= [r(t) + A(t,T)]\, dt + S(t,T)\, dW_t
\end{aligned}$$

and thus

$$dp = d(e^Y) = e^Y[dY + \frac{1}{2}d[Y]]$$

implies

$$\frac{dP(t,T)}{p(t,T)} = \left[r(t) + A(t,T) + \frac{1}{2}||S(t,T)||^2\right] dt + S(t,T)\, dW_t$$

as required.

□

Example 8.1.10 Synthetic Money Market Account
In a bond market, subject to the conditions enumerated before, it is possible to synthetically

create a locally risk–free bank account. This is accomplished by rolling over just maturing bonds.

Consider a portfolio V which, at any time, consists solely of bonds maturing at time $t+dt$. Suppose that there are n_t such bonds in the portfolio, so that

$$V_t = n_t p(t, t+dt)$$

By the self–financing condition,

$$\begin{aligned} dV_t &= n_t \; dp(t, t+dt) \\ &= n_t p(t, t+dt) \left\{ \left[r(t) + A(t, t+dt) + \frac{1}{2} ||S(t, t+dt)||^2 \right] \; dt + S(t, t+dt) \; dW_t \right\} \end{aligned}$$

Now as $dt \to 0$, also $A(t, t+dt) = -\int_t^{t+dt} \alpha(t,s) \; ds \to 0$, and $S(t, t+dt) = -\int_t^{t+dt} \sigma(t,s) \; ds \to 0$. Thus in the limit,

$$dV_t = r(t) V_t \; dt$$

which are just the dynamics of the MMA.

(Note, however, that the above argument is heuristic in nature: It requires, in any time interval, however short, the use of infinitely many types of securities.)

$\square$

In the riskneutral world, discounted bond price processes are martingales, and thus

$$\begin{aligned} p(t,T) &= \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r(s) \; ds} p(T,T) | \mathcal{F}_t \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r(s) \; ds} | \mathcal{F}_t \right] \end{aligned}$$

In a Brownian world, any equivalent measure is obtained from the objective measure by a Girsanov transformation — a consequence of the Martingale Representation Theorem. If

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{\int_0^T u \; dW_t - \frac{1}{2} \int_0^T ||u||^2 \; dt}$$

and if $L(t) = \mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{F}_t]$ is the associated likelihood process, then $dL_t = u_t L_t \; dW_t$. Now if $\hat{p}(t,T) = \frac{p(t,T)}{A_t}$, then (under $\mathbb{Q}$)

$$d\hat{p}(t,T) = \hat{p}(t,T) v(t,T) \; d\hat{W}_t$$

(where $\hat{W}_t$ is a $\mathbb{Q}$–Brownian motion), so that

$$dp(t,T) = r(t) p(t,T) \; dt + p(t,T) v(t,T) \; d\hat{W}_t$$

Hence under $\mathbb{P}$, we have dynamics

$$\frac{dp(t,T)}{p(t,T)} = [r(t) - u(t) v(t,T)] \; dt + v(t,T) \; dW_t$$

i.e. in a Brownian world bond price dynamics are necessarily of the form $dp = pm \; dt + pv \; dW_t$.

8.2 Modelling the Short Rate

Short rate models are bond market models where the only explanatory variable is the short rate r. This was the earliest approach to bond market models, dating back to the paper by vasiček (1977), but short rate models have limited power. Nevertheless, principal component analysis shows that typically $80-90\%$ of price variation in the bond market can be explained by a single factor, so these models are not wholly devoid of realism.

When we specify only the short rate, the *only* exogenously given asset is the MMA A_t. Zero coupon bonds will be regarded not as primitive securities, but as derivatives of the short rate.

Question: Are bond prices uniquely determined by the $\mathbb{P}$–dynamics of the short rate?

We assume that we live in a Brownian world governed by an objective probability measure $\mathbb{P}$, with change driven by a (multidimensional) Brownian motion W_t. We further assume short rate dynamics of the form

$$dr(t) = \mu(t,r)\,dt + \sigma(t,r)\,dW_t$$

i.e. the short rate is an Itô diffusion.

The answer to the above question is **No!**

- The above bond market is clearly incomplete:
  - We are able to execute trading strategies which consist of putting all our money in the bank account only. This clearly doesn't give us enough freedom to replicate all possible $\mathcal{F}_T$–measurable ransom variables.
  - There is at least one source of randomness, but there are no risky assets.
  - Under any measure , the discounted MMA $\frac{A_t}{A_t}$ is a martingale. hence any measure equivalent to $\mathbb{P}$, including $\mathbb{P}$ itself, is an EMM. The EMM is not unique.
- If $\mathbb{Q} \sim \mathbb{P}$ is any equivalent measure, then $\mathbb{Q}$ generates an arbitrage-free bond market with prices
$$p(t,T) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,ds} | \mathcal{F}_t\right]$$
- In the Black–Scholes model, we also had one source of uncertainty, but there option prices are determined by the dynamics of an underlying which is traded. The crucial difference here is that the underlying is the short rate, which is not a traded security.

Nevertheless, bonds of different maturities must satisfy certain internal consistency conditions in order to exclude arbitrage. For example, if $T_1 < T_2$, then $p(t,T_1) \geq p(t,T_2)$, or else there will be arbitrage (assuming positive rates).

If we have d sources of noise (i.e. W_t is a d–dimensional Brownian motion), then we may pick d maturities, and regard the bonds of those maturities as "primitive" securities; bonds of all other maturities will be "derivative". Our market now has as many risky primitive assets as securities, and is therefore complete.

8.2.1 The Term Structure PDE

Assume that we have an arbitrage–free bond market, with $\mathbb{P}$–short rate dynamics given by

$$dr(t) = \mu(t,r)\,dt + \sigma(t,r)\,dW_t$$

where W_t is a *one–dimensional* $\mathbb{P}$–Brownian motion. We restrict to one dimension purely for ease of exposition – similar results hold in the multidimensional case.

Also assume that the price of a T–bond at time t is given by a sufficiently smooth and regular function F:

$$p(t,T) = F(t, r(t); T) = F^T(t, r)$$

By taking two bonds of different maturities S and T, we are able to create a locally riskless portfolio. Arbitrage considerations then dictate that the drift of this portfolio is equal to the short rate. As usual, this yields a PDE, as we now show.

First note that by Itô's formula

$$dF^T = \left[F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T\right] dt + \sigma F_r^T \, dW_t$$

(where subscripts denote partial derivatives), so that

$$\frac{dF^T}{F^T} = \alpha^T \, dt + \sigma^T \, dW_t$$

where

$$\alpha^T(t,r) = \frac{F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T}{F^T}$$
$$\sigma^T(t,r) = \frac{\sigma F_r^T}{F^T}$$

Consider now a portfolio V consisting of S– and T–bonds with relative weights w^S, w^T respectively. Then

$$\frac{dV}{V} = w^S \frac{dF^S}{F^S} + w^T \frac{dF^T}{F^T}$$

To eliminate risk, set $w^S\sigma^S + w^T\sigma^T = 0$. Since weights add up to 1, we therefore obtain

$$w^S = \frac{\sigma^T}{\sigma^T - \sigma^S} \qquad w^T = -\frac{\sigma^S}{\sigma^T - \sigma^S}$$

Then

$$\frac{dV}{V} = \frac{\alpha^S\sigma^T - \alpha^T\sigma^S}{\sigma^T - \sigma^S} \, dt$$

i.e.

$$\frac{\alpha^S\sigma^T - \alpha^T\sigma^S}{\sigma^T - \sigma^S} = r$$

i.e.

$$\frac{\alpha^S(t,r) - r}{\sigma^S} = \frac{\alpha^T(t,r) - r}{\sigma^T}$$

Now α^T is just the drift of the bond price $p(t,T) = F^T(t,r)$, and σ^T is its volatility. Thus

$$\frac{\alpha^T(t,r) - r}{\sigma^T} = \text{ Market Price of Risk} = \lambda$$

i.e. all bonds have the same market price of risk $\lambda = \lambda(t,r)$. λ is independent of maturity (though it may vary over time).

Proposition 8.2.1 (Term Structure PDE)
In an arbitrage–free one–factor short rate model $dr = \mu\, dt + \sigma\, dW_t$ *there is a process* $\lambda(t,r)$ *such that*

$$\frac{\alpha^T(t,r) - r}{\sigma^T} = \text{Market Price of Risk} = \lambda$$

Hence all bonds satisfy the following PDE

$$\boxed{\begin{aligned} F_t^T + (\mu - \lambda\sigma)F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T &= 0 \\ F^T(T,r) &= 1 \end{aligned}}$$

Proof: Since

$$\begin{aligned} \alpha^T(t,r) &= \frac{F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T}{F^T} \\ \sigma^T(t,r) &= \frac{\sigma F_r^T}{F^T} \end{aligned}$$

we have

$$\frac{F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T}{\sigma F_r^T} = \lambda$$

which can easily be manipulated to yield the term structue PDE.

□

Using the Feynman–Kač formula, we see that the bond prices are given by

$$\begin{aligned} F^T(t,r) = F(t,r;T) &= \mathbb{E}_{\mathbb{Q}^\lambda}^{t,r}\left[e^{-\int_t^T r(s)\, ds}\right] \quad \text{i.e.} \\ p(t,T) &= \mathbb{E}_{\mathbb{Q}^\lambda}^{t,r}\left[e^{-\int_t^T r(s)\, ds} p(T,T)\right] \end{aligned}$$

where

$$\begin{aligned} dr &= (\mu - \sigma\lambda)\, ds + \sigma\, d\hat{W}_s \quad s \geq t \\ r(t) &= r \end{aligned}$$

are the dynamics of r under $\mathbb{Q}^\lambda$. Note that, since r_t is a Markov process, we have

$$\mathbb{E}_{\mathbb{Q}^\lambda}^{t,r}\left[e^{-\int_t^T r(s)\, ds} p(T,T)\right] = \mathbb{E}_{\mathbb{Q}^\lambda}\left[e^{-\int_t^T r(s)\, ds} p(T,T)|\mathcal{F}_t\right]$$

From the fact that

$$p(t,T) = \mathbb{E}_{\mathbb{Q}^\lambda}\left[e^{-\int_t^T r(s)\, ds} p(T,T)|\mathcal{F}_t\right]$$

it follows that each $\mathbb{Q}^\lambda$ is a risk–neutral measure (i.e. an EMM for the MMA).

We can also get the risk–neutral short rate dynamics from Girsanov's Theorem: a Girsanov transformation which effects the change of measure from real–world to risk–neutral has a Girsanov kernel equal to the negative of the market price of risk. Thus $-\sigma\lambda$ is added to the drift when we change the measure. Each market price of risk process λ gives a different risk–neutral measure $\mathbb{Q}^\lambda$.

To summarize:

- In an arbitrage–free short rate model, all bonds have the same market price of risk, regardless of maturity.

- Different market prices of risk yield different risk–neutral measures — The bond market is not complete.
- The agents in the market will (implicitly) determine λ and thus $\mathbb{Q}^\lambda$.

8.2.2 Martingale Models of the Short Rate

We model the short rate directly under a fixed riskneutral measure $\mathbb{Q}$. This is the EMM chosen by market participants, and should, in principle, be hidden in the term structure of bond prices. By calibrating a short rate model to bond prices, the market price of risk, and thus the market EMM, can be determined. This procedure is known as *inverting the yield curve*, and works as follows:

(1) Choose a short rate model (Ho–Lee, Vasiček, Cox–Ingersoll–Ross, Black–Derman–Toy) involving one or more parameters $\alpha = (\alpha_1, \ldots, \alpha_n)$. The $\mathbb{Q}$–dynamics of the short rate are given by

$$dr(t) = \mu(t, r(t); \alpha)\, dt + \sigma(t, r(t); \alpha) dW_t$$

(2) Solve the term structure PDE. In the risk–neutral world, the market price of risk is $\lambda = 0$, and thus the PDE is

$$\begin{aligned} F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T &= 0 \\ F^T(T, r) &= 1 \end{aligned}$$

for all maturities T. This yields theoretical bond prices

$$p(t, T; \alpha) = F^T(t, r_t; \alpha)$$

(3) Go to the market, and "observe" the empirical term structure of bond prices $\{p^*(0, T) : T \geq 0\}$.

(4) Choose α so that the theoretical prices $p(0, T; \alpha)$ fit the empirical prices $p^*(0, T)$ "as closely as possible" (where "close" must be defined somehow. For example, one method would be to pick maturities $T_1, \ldots, T_n$ and to pick $\alpha_1, \ldots, \alpha_n$ so that

$$\sum_{k=1}^{n} (p(0, T; \alpha) - p^*(0, T))^2$$

is minimized.) Let α^* be this "best" parameter.

(5) We now have dynamics

$$dr(t) = \mu(t, r(t); \alpha^*)\, dt + \sigma(t, r(t); \alpha^*)\, dW_t$$

under the risk–neutral measure. We can also, in principle, observe the real–world dynamics

$$dr(t) = \bar{\mu}\, dt + \bar{\sigma}\, d\bar{W}_t$$

Since $\mu = \bar{\mu} - \sigma\lambda$, we now know the market price of risk λ, and thus $\mathbb{Q} = \mathbb{Q}^\lambda$.

(6) Ideally, we would like to have

$$p(0, T; \alpha^*) = p^*(0, T) \qquad \text{for all } T$$

However, these are infinitely equations (one for each T), in only finitely many unknowns (the $\alpha_1, \ldots, \alpha_n$). This system is over–determined, and the model can not be made to fit the initial term structure of bond prices.

(7) However, if we choose α to be an infinite dimensional vector, rather than a finite dimensional one, there may be sufficient room to fit the term structure exactly. For example, the Ho–Lee model is given by

$$dr(t) = \theta(t)\, dt + \sigma\, dW_t$$

where σ is a constant, and W_t a one–dimensional Brownian motion. here $\alpha = (\theta(t) : t \geq 0)$ is an infinite –dimensional vector. The Ho–Lee model *can* be fitted to the empirically observed term structure, but this is not obvious *a priori.*

(8) Once we've parametrized our model,we can fit other interest rate derivatives.

8.2.3 Common Short Rate Models

The following are common short rate models with just one source of noise:

- Vasiček:
$$dr = (b - ar)\, dt + \sigma\, dW_t$$
where a, b, σ are constants.

- Cox–Ingersoll–Ross:
$$dr = (a - br)\, dt + \sigma\sqrt{r}\, dW_t$$
where a, b, σ are constants.

- Dothan or Rendlemann–Barter:
$$dr = ar\, dt + \sigma r\, dW_t$$
where a, σ are constants.

- Merton:
$$dr = a\, dt + \sigma\, dW_t$$
where a, σ are constants.

- Ho–Lee:
$$dr = \theta(t)dt + \sigma\, dW_t$$
where σ are constants.

- Hull–White (extended Vasiček):
$$dr = (b(t) - a(t)r)\, dt + \sigma(t)\, dW_t$$

- Hull–White (extended CIR):

$$dr = (b(t) - a(t)r)\ dt + \sigma(t)\sqrt{r}\ dW_t$$

- Black–Derman–Toy:

$$dr = a(t)r\ dt + \sigma(t)r\ dW_t$$

- Black–Karasinski:

$$dr = (a(t)r + b(t)r\ln r)\ dt + \sigma(t)r\ dW_t$$

All of the above can be written as

$$dr = (\alpha_1(t) + \alpha_2(t)r + \alpha_3(t)r\ln r)\ dt + (\beta_1(t) + \beta_2(t)r)^\nu\ dW_t$$

8.2.4 Term Structure Derivatives

Consider the general short rate model $dr(t) = \mu(t, r)\ dt + \sigma(t, r)\ dW_t$. Suppose that an interest rate derivative has a terminal payoff $\Phi(T, r_T)$ and a dividend rate $q(t, r_t)$ over the interval $[0, T]$. The time–t price of the derivative is obtained via an arbitrage argument: Start with a portfolio V consisting of one derivative F and $-n$ T–bonds p. Because of the dividends, we obtain

$$dV = dF - n\ dp + q\ dt$$

But choosing $n = \frac{\partial F}{\partial r} / \frac{\partial p}{\partial r}$ will make the portfolio locally riskless, and we obtain

$$\frac{F_t + \frac{1}{2}\sigma^2 F_{rr} - rF + q}{\frac{\partial F}{\partial r}} = \frac{P_t + \frac{1}{2}\sigma^2 P_{rr} - rP}{\frac{\partial P}{\partial r}}$$

Now the term structure PDE states that

$$P_t + \frac{1}{2}\sigma^2 P_{rr} - rP = -(\mu - \sigma\lambda)P_r$$

and thus

$$\begin{cases} F_t + (\mu - \sigma\lambda)F_r + \dfrac{1}{2}\sigma^2 F_{rr} - rF + q = 0 \\ \hfill F(t, r_T) = \Phi(T, r_T) \end{cases}$$

This is the generalized term structure equation for an interest rate derivative F (where $\lambda = 0$ if we model the short rate in the risk–neutral world).

The value of the interest rate derivative is clearly

$$F(t, r_t; T) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r_u\ du}\Phi(T, r_T) + \int_t^T e^{-\int_t^s r_u\ du} q(s, r_s)\ ds\right]$$

A trivial generalization of the Feynman–Kac argument shows that the solution of a PDE of the form

$$F_t + \mu F_x + \frac{1}{2}\sigma^2 F_{xx} - rF + h = 0 \qquad F(T, x) = \Phi(T, x)$$

is given by

$$F(t, x) = \mathbb{E}^{r,x}\left[\int_t^T e^{-\int_t^u r_s\ ds} h(u, X_u)\ du + e^{-\int_t^T r_s\ ds}\Phi(T, X_T)\right]$$

where

$$dX_s = \mu\ ds + \sigma\ dW_s \quad \text{for } t \le s \le T \qquad \text{and} \qquad X_t = x$$

Example 8.2.2 (a) A call with strike K and expiry τ on a discount bond $p(t,T)$ (with $T > \tau$) has $q = 0$ and $\Phi(\tau, r) = (p(\tau,T) - K)^+$. To calculate the option price, we first have to solve the term structure PDE to get the bond prices, and then once more to price the option.

(b) An interest swap (pay–fixed) can be idealized as a contract paying a divided rate $h(t, r_t) = r_t - r^*$, where r^* is the agreed-upon fixed rate (the swap rate at inception). Here $\Phi(t, r_T) = 0$, and so

$$F(t, r_t) = \mathbb{E}_{\mathbb{Q}}\left[\int_t^T e^{-\int_t^s r_u\,du}(r_s - r^*)\,ds\right]$$

Now a floating rate note paying a continuous rate r_t must be priced at par $= 1$ in order to avoid arbitrage (Why?), and thus

$$\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} + \int_t^T e^{-\int_t^s r_u\,du} r_s\,ds\right] = 1$$

It follows that

$$\mathbb{E}_{\mathbb{Q}}\left[\int_t^T e^{-\int_t^s r_u\,du} r_s\,ds\right] = 1 - \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] = 1 - p(t,T)$$

Hence the value of this idealized swap is

$$F(t, r_t) = 1 - p(t,T) - r^* \int_t^T p(t,s)\,ds$$

The swap rate at time t for maturity T sets the value of the swap to zero, and is

$$r^*(t,T) = \frac{1 - p(t,T)}{\int_t^T p(t,s)\,ds}$$

(c) A cap can be idealized as a derivative with zero terminal payoff and a dividend rate $q(t, r_t) = (r_t - \hat{r})^+$, where $\hat{r}$ is the cap rate.

□

We will spend quite a bit of effort pricing options on discount bonds in the next few pages. But what about coupon –bearing bonds, which are, after all, more commonly traded in the market? *Jamshidian's Trick* sometimes holds the answer: In a short rate model, a call on a coupon bearing bond can be priced as a portfolio of calls on zero coupon bonds, provided that the value $p(t,T) = p(t, r_t; T)$ of the zero coupon bonds is a strictly decreasing function of the short rate.

Theorem 8.2.3 *Let $\mathcal{C}^{K,\tau}(t, r_t)$ be the time–t value of a call on a coupon bond $\mathcal{B}$, where τ is the expiry of the call, and K the strike. Suppose that the coupon bond pays a coupon Y_i at date T_i, where $\tau < T_1 < \cdots < T_N$. Let*

$$K^i = p(\tau, r^*, T_i)$$

where r^* solves

$$\mathcal{B}(\tau, r^*) = K$$

Recall that $\mathcal{B}(\tau, r) = \sum_i Y_i p(\tau, r, T_i)$. Since each $p(t, r, T_i)$ is a decreasing function of r, so is $\mathcal{B}(t, r)$, which implies that r^* is unique. Solve numerically for r^*, e.g. via bisection method. *Then*

$$\mathcal{C}^{K,\tau}(t, r) = \sum_i Y_i C^{K_i,\tau,T_i}(t, r)$$

where $C^{K,T,S}$ is the time–t value of a strike K, expiry T call on a zero coupon bond $p(t, S)$ (with $S \geq T$).

Proof: The payoff of the call on $\mathcal{B}$ is

$$(\mathcal{B}(\tau, r_\tau) - K)^+ = \left(\sum_i Y_i p(\tau, r_\tau, T_i) - K \right)^+$$

Since each $p(t, r, T)$ is decreasing in r, so is $\mathcal{B}(\tau, r_\tau)$. Let r^* be the unique value of r_τ for which the call $\mathcal{C}$ expires at the money, i.e. for which

$$\mathcal{B}(\tau, r^*) = K$$

Now define $K_i = p(\tau, r^*, T_i)$. Then

$$\sum_i Y_i K_i = K$$

Now consider two cases:

Case 1: If $r_\tau < r^*$, then

$$\sum_i Y_i p(\tau, r_\tau, T_i) > \sum_i Y_i p(\tau, r^*, T_i) = K$$

and

$$p(\tau, r_\tau, T_i) > p(\tau, r^*, T_i) = K_i$$

Thus if $\mathcal{C}^{K,\tau}$ expires in the money, then so does each C^{K_i,τ,T_i}, and

$$\begin{aligned}
\left(\sum_i Y_i p(\tau, r_\tau, T_i) - K \right)^+ &= \sum_i Y_i p(\tau, r_\tau, T_i) - K \\
&= \sum_i Y_i \left(p(\tau, r_\tau, T_i) - K_i \right) \\
&= \sum_i Y_i \left(p(\tau, r_\tau, T_i) - K_i \right)^+
\end{aligned}$$

Case 2: If $r_\tau \geq r^*$, then $\sum_i Y_i p(\tau, r_\tau, T_i) \leq K$ and $p(\tau, r_\tau, T_i) \leq K_i$. Thus if $\mathcal{C}^{K,\tau}$ expires out of the money, then so does each C^{K_i,τ,T_i}, so

$$\begin{aligned}
0 &= \left(\sum_i Y_i p(\tau, r_\tau, T_i) - K \right)^+ \\
&= \sum_i Y_i \left(p(\tau, r_\tau, T_i) - K_i \right)^+
\end{aligned}$$

Hence in either case

$$\mathcal{C}^{K,\tau}(\tau, r_\tau) = \sum_i Y_i C^{K_i,\tau,T_i}(\tau, r_\tau)$$

Thus, by the law of one price,

$$\mathcal{C}^{K,t}(t, r_t) = \sum_i Y_i C^{K_i,t,T_i}(t, r_t)$$

for all $t \leq \tau$ as well.

$\dashv$

8.2.5 Lognormal Models

The Dothan, Rendleman–Barter, Black–Derman–Toy and Black–Karasinski all yield lognormal short rate dynamics. All suffer from the following problem: Let A_t denote the money market account, with $dA_t = r_t A_t\, dt, A_0 = 1$. Then

$$\mathbb{E}\left[e^{-\int_0^t r_s\, ds}\right] \approx \mathbb{E}\left[e^{\frac{t}{2}(r_0 + r_t)}\right]$$

for sufficiently small t. Now, since r_t is lognormal, define $Y_t = \ln r_t$. Then we have an expectation of the form

$$\mathbb{E}\left[e^{\frac{t}{2}e^{Y_t}}\right] = \mathbb{E}\left[e^{e^Z}\right]$$

for some normally distributed Z. Now

$$\int_{-\infty}^{\infty} e^{e^z} e^{-z^2/2}\, dz = \infty$$

as $e^{e^z} >> e^{-z^2/2}$ for reasonable values of z. Hence $\mathbb{E}[A(t)] = \infty$ even if t is small, i.e. the bank account, on average, explodes.

Indeed, it can be shown that

$$\mathbb{E}\left[\frac{1}{p(t,T)}\right] = \infty \qquad \text{for all } t > 0$$

One consequence of this *lognormal explosion* is that one cannot price *Eurodollar futures.*

8.3 Affine Term Structure Models

8.3.1 Mechanics of ATS models

Definition 8.3.1 A short rate model is said to possess *affine term structure* (ATS) if bond prices are given by

$$p(t,T) = F^T(t, r(t)) = e^{A(t,T) - B(t,T) r(t)}$$

where $A(t,T), B(t,T)$ are (sufficiently regular) deterministic functions.

$\square$

Note that not all short rate models are affine term structure models. However, the class of affine term structure models is quite well understood: They are those for which both the drift and the volatility–squared are affine functions of the short rate.

For consider a short rate model with *risk–neutral* dynamics $dr(t) = \mu(t,r)\ dt + \sigma(t,r)\ dW_t$ and suppose that bond prices are of the form $p(t,T) = F^T(t,r(t)) = e^{A(t,T)-B(t,T)r(t)}$. Substituting this expression into the term structure PDE

$$\begin{aligned} F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T = 0 \\ F^T(T,r) = 1 \end{aligned}$$

we obtain

$$A_t - \mu B + \frac{1}{2}\sigma^2 B^2 - (1+B_t)r = 0$$

Moreover, since $p(T,T) = 1$, we must have $A(T,T) = 0 = B(T,T)$.

If we assume that the drift and volatility of the short rate can be expressed in the form

$$\begin{aligned} \mu(t,r) &= \alpha(t)r + \beta(t) \\ \sigma^2(t,r) &= \gamma(t)r + \delta(t) \end{aligned}$$

then we obtain

$$\left[A_t - \beta B + \frac{1}{2}\delta B^2\right] = \left[1 + B_t + \alpha B - \frac{1}{2}\gamma B^2\right] r$$

The lefthand side is independent of r, whereas the righthand side contains r. This can happen only if both sides are identically zero, so that we obtain a coupled system of differential equations:

$$\begin{cases} A_t(t,T) = \beta(t)B(t,T) - \frac{1}{2}\delta(t)B^2(t,T) \\ A(T,T) = 0 \end{cases}$$

$$\begin{cases} B_t(t,T) = -\alpha(t)B(t,T) + \frac{1}{2}\gamma(t)B^2(t,T) - 1 \\ B(T,T) = 0 \end{cases}$$

Note that the bottom equation (a *Riccatti equation*) does not contain A, and can therefore be solved (in principle, although this may be quite hard). The solution can then be plugged into the top equation to solve for A. To solve this equation, simply integrate both sides (from t to T).

Thus a short rate model has affine term structure whenever μ, σ are of the form $\mu(t,r) = \alpha(t)r + \beta(t)$ and $\sigma^2(t,r) = \gamma(t)r + \delta(t)$. The Ho–Lee, Cox–Ingersoll–Ross, Merton, Vasiček and Hull–White models all have ATS. The Dothan and Black–Derman–Toy models do not.

Example 8.3.2 <u>The Vasiček Model</u>
Here we have $dr_t = (b - ar)\ dt + \sigma\ dW_t$, where a, b, σ are constants. Thus we have $\alpha = -a, \beta = b, \gamma = 0, \delta = \sigma^2$, all constant.
The system of differential equations that must be solved is therefore

$$\begin{cases} A_t = bB - \frac{1}{2}\sigma^2 B^2 \\ A(T,T) = 0 \end{cases}$$

$$\begin{cases} B_t = aB - 1 \\ B(T,T) = 0 \end{cases}$$

The bottom equation is a first order linear equation. This can easily be solved: Use e^{-at} as an integrating factor to obtain

$$B(t,T) = e^{at}\left[\frac{1}{a}e^{-at} + C(T)\right]$$

and then use $B(T,T) = 0$ to get $C(T) = -\frac{1}{a}e^{-aT}$. Hence

$$B(t,T) = \frac{1}{a}\left[1 - e^{-a(T-t)}\right]$$

Plug this into the equation for A to obtain

$$\begin{cases} A_t = bB(t,T) - \frac{1}{2}\sigma^2 B^2(t,T) \\ \quad = b\frac{1}{a}\left[1 - e^{-a(T-t)}\right] - \frac{1}{2}\sigma^2\frac{1}{a^2}\left[1 - e^{-a(T-t)}\right]^2 \\ A(T,T) = 0 \end{cases}$$

Integrate both sides:

$$\begin{aligned} A(t,T) &= A(T,T) - \int_t^T A_t(s,T)\,ds \\ &= \frac{1}{2}\sigma^2\int_t^T \frac{1}{a^2}\left[1 - e^{-a(T-s)}\right]^2\,ds - b\int_t^T \frac{1}{a}\left[1 - e^{-a(T-s)}\right]\,ds \\ &= -\frac{\sigma^2 B^2}{4a} + \frac{(B - (T-t))(ab - \frac{1}{2}\sigma^2)}{a^2} \end{aligned}$$

Now that A, B have been found, bond prices are given by the equation $p(t,T) = e^{A(t,T)-B(t,T)r(t)}$.

□

In order to invert the yield curve in the above example, the parameters $a, b\sigma$ must now be chosen so that the model fits empirical (observed) term structure of bond prices $\{p^*(0,T) : T \geq 0\}$ as "closely" as possible. Clearly, however, we have infinitely many bond prices, but only three parameters, i.e. the system is highly over-determined, and therefore we cannot generally choose a, b, σ such that $e^{A(0,T)-B(0,T)r_0} = p^*(0,T)$, i.e. the model cannot be made to fit the observed term structure exactly (unless we are astoundingly fortunate). The Vasiček model is able to fit, exactly, just 3 bonds.

Example 8.3.3 Cox–Ingersoll–Ross model The risk–neutral short rate dynamics assumed are

$$dr_t = (b - ar_t)\,dt + \sigma\sqrt{r_t}\,dW_t, \qquad a, b, \sigma, r_0 > 0$$

This is mean reverting (to b/a). Since the volatility term $\sigma\sqrt{r_t}$ tends to zero as $r_t \to 0$ (which is consistent with observation), positive rates are assured (which is also consistent with

observation). Postulating $p(t,T) = e^{A(t,T)-B(t,T)r_t}$, we quickly determine, by substituting into the term structure PDE, that

$$\begin{aligned} B_t &= aB + \frac{1}{2}\sigma^2 B^2 - 1 \qquad && B(T,T) = 0 \\ A_t &= bB && A(T,T) = 0 \end{aligned}$$

To solve the Riccati equation for B, we try a solution of the form

$$B(t,T) = \frac{X(t)}{cX(t)+d}$$

Then

$$B_t = \frac{X_t}{cX+d} - \frac{cXX_t}{(cX+d)^2}$$

and hence, substituting into the equation for B_t, we see that

$$-dX_t + X^2(ac + \frac{1}{2}\sigma^2 - c^2) + X(ad - 2cd) - d^2 = 0 \qquad X(T) = 0$$

Choose c to ensure that $a + \frac{1}{2}\sigma^2 - c^2 = 0$, i.e. $c = \frac{1}{2}(a + \sqrt{a^2 + 2\sigma^2})$. We then have a order linear differential equation

$$X_t + \kappa X = -d \qquad \text{where } \kappa = -a + 2c = \sqrt{a^2 + 2\sigma^2}$$

Since $X(T) = 0$,we see that

$$X(t) = \frac{d}{\kappa}[e^{\kappa(T-t)} - 1]$$

Hence

$$\begin{aligned} B(t,T) &= \frac{X(t)}{cX(t)+d} \\ &= \frac{e^{\kappa(T-t)} - 1}{\frac{1}{2}(\kappa + a)(e^{\kappa(T-t)} - 1) + \kappa} \\ &= \frac{2(e^{\kappa(T-t)} - 1)}{2\kappa + (a+\kappa)(e^{\kappa(T-t)} - 1)} \qquad \text{where } \kappa = \sqrt{a^2 + 2\sigma^2} \end{aligned}$$

Then $A(t,T)$ is obtained by integrating:

$$A(t,T) = A(T,T) - \int_t^T A_t(s,T)\, ds = -b \int_t^T B(s,T)\, ds$$

The solution is

$$A(t,T) = \frac{2b}{\sigma^2} \ln \left[\frac{2\kappa e^{\frac{1}{2}(a+\kappa)(T-t)}}{2\kappa + (a+\kappa)(e^{\kappa(T-t)} - 1)} \right]$$

as can be verified by differentiation.

$\square$

Example 8.3.4 Ho–Lee Model

We are given risk–neutral short rate dynamics $dr(t) = \theta(t)\ dt + \sigma\ dW_t$, where $\theta(t)$ is deterministic and σ a constant. The model has ATS with $\alpha = 0, \beta = \theta, \gamma = 0, \delta = \sigma^2$. This leads to two differential equations. The first is

$$\begin{cases} B_t = -1 \\ B(T,T) = 0 \end{cases}$$

which has solution $B(t,T) = T - t$ (as can be seen by integrating both sides from t to T). The second DE is

$$\begin{cases} A_t = \theta(t)B(t,T) - \dfrac{1}{2}\sigma^2 B^2(t,T) \\ \quad\;\, = \theta(t)(T-t) - \dfrac{1}{2}\sigma^2(T-t)^2 \\ A(T,T) = 0 \end{cases}$$

Integrating both sides from t to T yields

$$A(t,T) = -\int_t^T \theta(s)(T-s)\ ds + \frac{1}{6}\sigma^2(T-t)^3$$

We now choose the function $\theta(t)$ so as to fit the initial term structure of bond prices $\{p^*(0,T) : T \geq 0\}$, or, equivalently, the observed term structure of (instantaneous) forward rates $\{f^*(0,T) : T \geq 0\}$.

Recall that $f^*(0,T) = -\frac{\partial \ln p^*(0,T)}{\partial T}$. With affine term structure, we have $p^*(0,T) = e^{A(0,T)-B(0,T)r_0}$, so

$$\ln p^*(0,T) = -\int_0^T \theta(s)(T-s)\ ds + \frac{1}{6}\sigma^2 T^3 - r_0 T$$

Differentiating with respect to T, we see that

$$f^*(0,T) = \int_0^T \theta(s)\ ds - \frac{1}{2}\sigma^2 T^2 + r_0$$

Differentiating once more with respect to T, we obtain

$$\frac{\partial f^*(0,T)}{\partial T} = \theta(T) - \sigma^2 T$$

and thus we have found θ:

$$\theta(t) = f_T^*(0,t) + \sigma^2 t$$

We can use this to calculate $A(t,T)$:

$$\begin{aligned} A(t,T) &= \int_t^T (f_T^*(0,s) + \sigma^2 s)(s-T)\ ds + \frac{1}{6}\sigma^2(T-t)^3 \\ &= f^*(0,s)(s-T)|_t^T - \int_t^T f^*(0,s)\ ds + \sigma^2\left[\frac{s^3}{3} - \frac{s^2T}{2}\right]_t^T + \frac{1}{6}\sigma^2(T-t)^3 \\ &= f^*(0,t)(T-t) + \int_t^T \frac{\partial \ln p^*(0,s)}{\partial T}\ ds - \frac{1}{2}\sigma^2 t(T-t)^2 \\ &= f^*(0,t)(T-t) + \ln\left(\frac{p^*(0,T)}{p^*(0,t)}\right) - \frac{1}{2}\sigma^2 t(T-t)^2 \end{aligned}$$

Using the fact that the Ho–Lee model has ATS, we see that bond prices are given by

$$p(t,T) = \frac{p^*(0,T)}{p^*(0,t)} \exp\left(f^*(0,t)(T-t) - \frac{1}{2}\sigma^2 t(T-t)^2 - (T-t)r(t) \right)$$

where

$$\begin{aligned} r(t) &= r_0 + \int_0^t \theta(s)\, ds + \int_0^t \sigma\, dW_s \\ &= r_0 + f^*(0,t) - f^*(0,0) + \frac{1}{2}\sigma^2 t^2 + \sigma W_t \\ &= f^*(0,t) + \frac{1}{2}\sigma^2 t^2 + \sigma W_t \end{aligned}$$

because $r_0 = f^*(0,0)$. It follows that $\mathbb{E}[r_t] \to \infty$ (under the riskneutral measure). This is clearly a flaw in the model.

Since the short rate is Gaussian, future bond prices are lognormally distributed under the risk–neutral measure. In particular, there is a non–zero probability that a bond will, at some future date, trade above par (i.e. that interest rates become negative). This is clearly another flaw in the model.

Now that we've calculated the evolution of future bond prices and rates, let's have a look at future forward rates. Since $f(t,T) = -\frac{\partial \ln p(t,T)}{\partial T}$, we see that

$$\begin{aligned} f(t,T) &= f^*(0,T) - f^*(0,t) + \sigma^2 t(T-t) + r_t \\ &= f^*(0,T) + \sigma^2 t(T - \frac{1}{2}t) + \sigma W_t \end{aligned}$$

using the expression for r_t obtained earlier. Note that $f(t,t) = r_t$.

Now if we fix $t > 0$, we see that $\mathbb{E}[f(t,T)] \to \infty$ as $T \to \infty$. Indeed, for large values of T, $f(t,T) \approx kT$. Thus even if the initial forward curve is bounded above, it will be unbounded an instant later. This is another flaw in the Ho–Lee model.

□

Example 8.3.5 The Hull–White (extended Vasiček) Model

Consider the short rate model with risk–neutral dynamics

$$dr_t = (b(t) - ar_t)\, dt + \sigma\, dW_t$$

where $b(t)$ is deterministic, a, σ are constants and W_t is a one–dimensional Brownian motion. This is clearly an affine term structure model $dr_t = (\alpha(t)r_t + \beta(t))\, dt + \sqrt{\gamma(t)r_t + \delta(t)}\, dW_t$, with $\alpha(t) = -a, \beta(t) = b, \gamma(t) = 0$ and $\delta(t) = \sigma^2$. Substituting $p(t,T) = e^{A(t,T)-B(t,T)r_t}$ into the term structure PDE yields

$$\begin{aligned} &B_t(t,T) = aB(t,T) - 1 \\ &B(T,T) = 0 \end{aligned}$$

and

$$\begin{aligned} &A_t(t,T) = b(t)B(t,T) - \frac{1}{2}\sigma^2 B^2(t,T) \\ &A(T,T) = 0 \end{aligned}$$

Hence

$$B(t,T) = \frac{1}{a}(1 - e^{-a(T-t)})$$
$$A(t,T) = \int_t^T -b(u)B(u,T) + \frac{1}{2}\sigma^2 B^2(u,T)\, du$$

Fitting the initial term structure of bond prices is equivalent to fitting the initial term structure of forward rates. The latter is more convenient. Now since $f(0,T) = -\frac{\partial \ln P(0,T)}{\partial T} = -A_T(0,T) + B_T(0,T)r_0$, and since $B_T(t,T) = e^{-a(T-t)}$, we observe

$$\begin{aligned} f(0,T) &= \int_0^T b(u)B_T(u,T) + \sigma^2 B_T(u,T)B(u,T)\, du + B_T(0,T)r_0 \\ &= \int_0^T b(u)e^{-a(T-u)}\, du - \frac{\sigma^2}{2a^2}(1 - e^{-a(T-t)})^2 + e^{-aT}r_0 \end{aligned}$$

We side–step the computation of these integrals using the following trick: Define

$$x(t) = e^{-at}r_0 + \int_0^t b(u)e^{-a(t-u)}\, du$$
$$y(t) = \frac{\sigma^2}{2a^2}(1 - e^{-at})^2$$

Note that

$$\begin{aligned} x'(T) &= -ar_0 e^{-aT} + b(T) - a\int_0^T b(u)e^{-a(T-u)}\, du \\ &= -ax(T) + b(T) \end{aligned}$$

Now $f(0,T) = x(T) - y(T)$, and so

$$\begin{aligned} b(T) &= x' + ax \\ &= f_T(0,T) + y'(T) + ax(T) \\ &= f_T(0,T) + y'(T) = a[f(0,T) + y(T)] \end{aligned}$$

Thus, noting that $y(t) = \frac{\sigma^2}{2a^2}(1-e^{-at})^2 = \frac{1}{2}\sigma^2 B^2(0,t)$ and thus that $y'(t) = \frac{\sigma^2}{a}(1-e^{-at})e^{-at} = \sigma^2 B(0,t)B_T(0,t)$, we obtain

$$\begin{aligned} b(t) &= f_T^*(0,t) + \sigma^2 B(0,t)B_T(0,t) + a[f^*(0,t) + \frac{1}{2}\sigma^2 B^2(0,t)] \\ &= f_T^*(0,t) + af^*(0,t) + \frac{\sigma^2}{2a}[1 - e^{-2at}] \end{aligned}$$

This is the function $b(t)$ which will fit forward rates to the observed term structure $\{f^*(0,T) : T \geq 0\}$.

Since we now know $b(t)$ we can calculate $A(t,T)$:

$$A(t,T) = \int_t^T -b(u)B(u,T) + \frac{1}{2}\sigma^2 B^2(u,T)\, du$$

Now note that $b(u) = x'(u) + ax(u) = e^{-au}\frac{d\, xe^{au}}{du}$ so that

$$\begin{aligned}
\int_t^T b(u)B(u,T)\,du &= \frac{1}{a}\int_t^T e^{-au}\frac{d\,xe^{au}}{du}(1-e^{-a(T-u)})\,du \\
&= \frac{1}{a}\int_t^T (e^{-au}-e^{-aT})\,d(xe^{au}) \\
&= \frac{1}{a}\left[x(u)e^{au}(e^{-au}-e^{-aT})\right]_t^T - \frac{1}{a}\int_t^T xe^{au}\cdot -ae^{-au}\,du \\
&= -\frac{1}{a}x(t)(1-e^{-a(T-t)}) + \int_t^T x(u)\,du \\
&= -\left[f(0,t)+\frac{\sigma^2}{2}B^2(0,t)\right]B(t,T) + \int_t^T -\frac{\partial p(0,u)}{\partial T} + \frac{\sigma^2}{2}B^2(0,u)\,du \\
&= -f(0,t)B(t,T) - \ln\frac{p(0,T)}{p(0,t)} - \frac{\sigma^2}{2}B^2(0,t)B(t,T) + \int_t^T \frac{\sigma^2}{2}B^2(0,u)\,du
\end{aligned}$$

Hence

$$A(t,T) = f(0,t)B(t,T) + \ln\frac{p(0,T)}{p(0,t)} + \frac{\sigma^2}{2}\left[\int_t^T B^2(u,T) - B^2(0,u)\,du + B^2(0,t)B(t,T)\right]$$

Now, after a few lines of manipulation,

$$\int_t^T B^2(u,T) - B^2(0,u)\,du + B^2(0,t)B(t,T) = -\frac{1}{2a}B^2(t,T)(1-e^{-2at})$$

as you can easily check, substituting $B(t,T) = \frac{1}{a}(1-e^{-a(T-t)})$. Thus

$$A(t,T) = f(0,t)B(t,T) + \ln\frac{p(0,T)}{p(0,t)} - \frac{\sigma^2}{4a}B^2(t,T)(1-e^{-2at})$$

Substituting $p(t,T) = e^{A(t,T)-B(t,T)r_t}$ we obtain:

$$p(t,T) = \frac{p(0,T)}{p(0,t)}e^{f(0,t)B(t,T)-\frac{\sigma^2}{4a}B^2(t,T)(1-e^{-2at})-B(t,T)r_t}$$

□

We have thus found the following bond prices:

Theorem 8.3.6 *(a) In the Ho–Lee model, bond prices (fitted to the initial term structure) are given by*

$$p(t,T) = \frac{p(0,T)}{p(0,t)}\exp\left(f(0,t)(T-t) - \frac{1}{2}\sigma^2 t(T-t)^2 - (T-t)r(t)\right)$$

(b) In the Hull–White (extended Vasiček) model, bond prices (fitted to the initial term structure) are given by

$$p(t,T) = \frac{p(0,T)}{p(0,t)}\exp\left(f(0,t)B(t,T) - \frac{\sigma^2}{4a}B^2(t,T)(1-e^{-2at}) - B(t,T)r_t\right)$$

where $B(t,T) = \frac{1}{a}(1-e^{-a(T-t)})$.

□

8.3.2 Bond Options

In the chapter on changes of numéraire, we obtained the following general option formula: The price of a call C with strike K and maturity T on an underlying S is given by

$$C_0 = S_0 \mathbb{Q}_S(S_T \geq K) - Kp(0,T)\mathbb{Q}^T(S_T \geq K)$$

where $\mathbb{Q}_S, \mathbb{Q}^T$ are the EMM's associated with numéraires $S_t, p(t,T)$ respectively.

In order to use this formula, and to get Black–Scholes type solutions to option pricing problems, we assumed that the volatility of the securities is deterministic, and then obtained

Theorem 8.3.7 *If $\hat{S}_t = \frac{S_t}{p(t,T)}$ is an Itô process of the form $\frac{d\hat{S}_t}{\hat{S}_t} = \mu(t)\,dt + \sigma(t)\cdot\,dW_t$, and if $\sigma(t)$ is* deterministic*, then the value of a call C with maturity strike K and T on underlying security S is given by*

$$C_0 = S_0 N(d_1) - Kp(0,T)N(d_2)$$

where

$$\sigma_{av} = \sqrt{\frac{1}{T}\int_0^T ||\sigma(t)||^2\,dt}$$

$$d_1 = \frac{\ln\frac{S_0}{Kp(0,T)} + \frac{1}{2}\sigma_{av}^2 T}{\sigma_{av}\sqrt{T}} \qquad d_2 \quad = d_1 - \sigma_{av}\sqrt{T}$$

Put–call parity yields

$$P_0 = -S_0 N(-d_1) + Kp(0,T)N(-d_2)$$

for the price of a corresponding put.

□

We can now use this theorem to price bond options.

Example 8.3.8 Bond Options in the Ho–Lee Model
Consider a European call option C with strike K and maturity T on a discount bond $p(t,S)$ (where $S > T$). In the Ho–Lee model, with risk–neutral dynamics $dr_t = \theta(t)\,dt + \sigma\,dW_t$, bond prices have dynamics

$$\frac{dp(t,T)}{p(t,T)} = r\,dt - \sigma(T-t)\,dW_t$$

The drift term is r, because bond prices have drift r under the risk–neutral measure, just like all other traded securities. The volatility is obtained from the affine term structure: $p(t,T) = e^{A(t,T)-B(t,T)r_t}$, and we found that $B(t,T) = T - t$ (and we don't care about the value of $A(t,T)$ right now.) Thus the bond volatilities are deterministic: $p(t,T)$ has volatility $-\sigma(T-t)$ and $p(t,S)$ has volatility $-\sigma(S-t)$. Now the underlying security is $p(t,S)$, and $\hat{p}(t,S) = \frac{p(t,S)}{p(t,T)}$ has deterministic (indeed, *constant*) volatility $-\sigma(S-t)+\sigma(T-t) = -\sigma(S-T)$. This is because the volatility of a ratio of two assets is just the difference of their volatilities.

It follows that $\hat{p}(t,S)$ is lognormally distributed, and that $\ln\hat{p}(t,S)$ has variance $\sigma_{av}^2 T = \int_0^T \sigma^2(T-S)^2\,dt = \sigma^2(S-T)^2 T$. It follows that the price of the call is

$$C_0 = p(0,S)N(d_1) - Kp(0,T)N(d_2)$$

where

$$d_1 = \frac{\ln \frac{p(0,S)}{Kp(0,T)} + \frac{1}{2}\sigma^2(S-T)^2 T}{\sigma(S-T)\sqrt{T}}$$
$$d_2 = d_1 - \sigma(S-T)\sqrt{T}$$

□

Example 8.3.9 Bond Options in the Hull–White (extended Vasiček) Model
We tackle once more the problem of pricing a call with strike K and maturity T on a zero coupon bond $p(0,S)$, where $S > T$. It ought to be clear from the analysis of bond options in the Ho–Lee model that we need mainly to find the volatility of the bonds $p(t,T)$. Now, as for the Ho–Lee model, the riskneutral dynamics of $p(t,T)$ are

$$\frac{dp(t,T)}{p(t,T)} = r\ dt - B(t,T)\sigma\ dW_t$$

so that the volatility of $p(t,T)$ is $-\frac{\sigma}{a}(1 - e^{-a(T-t)})$. The asset ratio $\hat{p}_t = p(t,S)/p(t,T)$ therefore has volatility $\frac{\sigma}{a}(e^{-aS} - e^{-aT})e^{at}$ at time t. Thus the average volatility–squared is

$$\sigma_{av}^2 T = \int_0^T \frac{\sigma^2}{a^2}(e^{-aS} - e^{-aT})^2 e^{2at}\ dt = \frac{\sigma^2}{2a^3}(1 - e^{-a(S-T)})^2(1 - e^{-2aT})$$

We now find that the value of the call is simply

$$p(0,S)N(d_1) - Kp(0,T)N(d_2)$$

where

$$d_1 = \frac{\ln \frac{p(0,S)}{Kp(0,T)} + \frac{1}{2}\sigma_{av}^2 T}{\sigma_{av}\sqrt{T}} \qquad d_2 = d_1 - \sigma_{av}\sqrt{T}$$

□

8.4 The Heath–Jarrow–Morton Framework

8.4.1 The Set–Up

Up till now, we have studied interest rate models in which the short rate is the only explanatory variable. Such an approach has many obvious advantages:

- Specifying r as the solution of an SDE allows us to use Markov theory, which leads to PDE's (e.g., via the Feynman–Kač theorem, or the Kolmogorov forward and backward equations) that can be solved;
- If we're lucky, we can obtain analytical formulas for bond prices and bond option prices (as we did for the Ho–Lee and Hull–White (extended Vasiček) models.

However, the short rate modelling approach has some obvious disadvantages as well:

- It is unreasonable to regard the short rate as the only explanatory variable — it is difficult to incorporate views about different times in the future;

- It can be quite difficult to fit a realistic volatility structure;
- In order for the model to have even a remote chance of being correct, it is necessary to invert the yield curve (i.e. to fit the model to the initial term structure of bond prices). This can be quite difficult as well.

The Heath–Jarrow–Morton (HJM) approach circumvents some of these difficulties by specifying dynamics for the entire (uncountable) family of forward rates. For a fixed $T \geq 0$, assume that the forward rate $f(t,T)$ has "real–world" dynamics

$$df(t,T) = \alpha(t,T)\ dt + \sigma(t,T)\ d\hat{W}_t \qquad T \geq 0, 0 \leq t \leq T$$

where $\hat{W}_t$ is a *finite–dimensional* Brownian motion under the real world measure $\mathbb{P}$, and $\alpha(t,T)$ and $\sigma(t,T)$ are adapted (and sufficiently regular to ensure that most of the operations below are permissible. For example, it is often necessary to assume that $\alpha(t,T)$ is jointly measurable in the t– and T–variables.)

Thus we have infinitely many SDE's, one for each maturity T. Each such SDE has an initial condition, namely $f(0,T) = f^*(0,T)$, where $f^*(0,T)$ is the observed term structure. the advantage of this approach is that the initial term structure is fitted automatically — it is an initial condition! — so that inverting the yield curve becomes unnecessary. It is also easier to incorporate views about different maturities, because we have many different SDE's. (The disadvantage, of course, is that we have many, many SDE's.) These are still manageable, because we assume that the bond market is driven by finitely many sources of noise. But this leads to another difficulty:

Remarks 8.4.1 Given $\alpha(t,T), \sigma(t,T)$ and $\{f^*(0,T) : T \geq 0\}$, we can solve the SDE's for the forward rate, so that we have specified the entire term structure $\{f^*(t,T) : T \geq 0, 0 \leq t \leq T\}$ at all times and all maturities, and thus the entire term structure of bond prices

$$p(t,T) = e^{-\int_t^T f(t,u)\ du}$$

Since we have only finitely many sources of noise, and infinitely many traded assets, there is a possibility of arbitrage in the bond market, unless the bond prices are inter–related in a specific way (which amounts to all bond prices having the same market price of risk, for all source of noise). This will impose conditions on the functions α and σ.

□

Remarks 8.4.2 HJM is not a model, but a framework of models for the bond market; short rate models are another such framework. But whereas short rate models are generally Itô diffusions, and thus Markov processes, we can easily let α and σ depend on past history. HJM models therefore need not be Markov models.(Of course, short rate models do not really need to be Markov either, but then their dynamics cannot be given by diffusions. We shall discuss the relationship between short rate and HJM models in the next section.)

□

For a market model (driven by Brownian motions) with only finitely many securities, we know that the model is arbitrage–free if and only if we can construct a risk–neutral measure, and complete if that measure is unique. Equivalently, the market is complete if an only if there

are as many traded risky securities as Brownian motions, subject to some conditions which ensure that the traded securities, are, in some sense , independent (where "independent" is meant in the sense of linear algebra, and not probability). The fact that there are only finitely many sources of noise, but infinitely many traded assets, means that the market is "over–complete", i.e. that there may be many ways of replicating a security. Unless all such replicating portfolios have the same price, there will be arbitrage. In practice, all securities must have the same market price of risk. If that's the case, we can construct a riskneutral measure (via a Girsanov transformation), which implies that the market is arbitrage–free. The arbitrage theory we've developed thus far only applies to markets with just finitely many traded securities, and it isn't at all clear that the impossibility of arbitrage implies the existence of a riskneutral measure (i.e. a measure under which all uncountably many zero coupon bond prices, when discounted, become martingales). We *can*, however, construct riskneutral measures for any finite subset of zero coupon bonds. Nevertheless, it is highly desirable to have a single riskneutral measure for all bonds simultaneously (because prices of securities are then just expected discounted payoffs, where the expectation is taken w.r.t. the riskneutral measure). We will therefore try to impose a strong form of the no–arbitrage condition: The existence of a riskneutral measure for all bonds.

To enable us to construct such a riskneutral measure, there must be relationships between $\alpha(t,T)$ and $\sigma(t,T)$ that must hold if the HJM model is to be arbitrage—free:

Proposition 8.4.3 *Assume that the bond market is arbitrage–free in the strong sense, i.e. assume that there is a risk–neutral measure for bonds of all maturities. Then there is a (multidimensional) process $\lambda(t)$ such that, for all maturities T,*

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)^{tr}\, ds + \sigma(t,T)\lambda(t)$$

Proof: Recall that

$$\frac{dp(t,T)}{p(t,T)} = \left[r(t) + A(t,T) + \frac{1}{2}||S(t,T)||^2\right]\, dt + S(t,T)\, d\hat{W}_t$$

where $\hat{W}_t$ is a $\mathbb{P}$–Brownian motion. Here

$$A(t,T) = -\int_t^T \alpha(t,s)\, ds$$
$$S(t,T) = -\int_t^T \sigma(t,s)\, ds$$

If we use a Girsanov transformation with kernel $-\lambda$ to change to a new measure $\mathbb{Q}$, then new dynamics of $p(t,T)$ are

$$\frac{dp(t,T)}{p(t,T)} = \left[r(t) + A(t,T) + \frac{1}{2}||S(t,T)||^2 - S(t,T)\lambda(t)\right]\, dt + S(t,T)\, dW_t$$

where W_t is a $\mathbb{Q}$–Brownian motion. For $\mathbb{Q}$ to be a riskneutral measure, each $p(t,T)$ must have drift $r(t)$, i.e.

$$A(t,T) + \frac{1}{2}||S(t,T)||^2 - \sigma(t,T)\lambda(t) = 0$$

This shows that λ is just the market price of risk of $p(t,T)$ at time t, for all T: All bonds have the same market price of risk.

Differentiating this equation with respect to T yields

$$-\alpha(t,T) + \sigma(t,T)\int_t^T \sigma(t,s)^{tr}\, ds + \sigma(t,T)\lambda(t) = 0$$

□

Suppose that we have an HJM model driven by d sources of noise, so that each $\sigma(t,T)$ is a d–dimensional row vector $\sigma = (\sigma_1, \ldots, \sigma_d)$, and $\lambda = (\lambda_1, \ldots, \lambda_d)^{tr}$ is a d–dimensional column vector. We then have

$$\alpha(t,T) = \sum_{i=1}^{d} \sigma_i(t,T)\int_t^T \sigma_i(t,s)\, ds + \sum_{i=1}^{d}\sigma_i(t,T)\lambda_i(t) \tag{*}$$

If we take α and σ as given, we can try and solve for λ. We then have uncountably many equations in just d unknowns $\lambda_1(t), \ldots, \lambda_d(t)$ — one equation for each T. Thus α, σ cannot be specified arbitrarily. What we can do is

- Specify the volatility surface $\sigma(t,T)$.
- Choose d benchmark maturities $T_1, \ldots, T_d$ and specify $\alpha(t,T_1), \ldots, \alpha(t,T_d)$.
- Solve the system $(*)$ of d equations for the d unknowns $\lambda_1(t), \ldots, \lambda_d(t)$.

All the other $\alpha(t,T)$ (for $T \neq$ a bench mark maturity) are now given by $(*)$.

8.4.2 Martingale Modelling

As for short rate models, it is often convenient to bypass the necessity of estimating the market price of risk, and to model directly under the risk–neutral measure $\mathbb{Q}$. i.e. we write

$$df(t,T) = \alpha(t,T)\, dt + \sigma(t,T)\, dW_t \qquad T \geq 0, 0 \leq t \leq T f(0,T) \quad = f^*(0,T)$$

where W_t is a $\mathbb{Q}$–Brownian motion. Under $\mathbb{Q}$, the market price of risk is $\lambda = 0$, so we obtain:

Proposition 8.4.4 (HJM Drift Conditions)
The riskneutral dynamics of forward rates satisfy the following conditions:

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)^{tr}\, ds$$

□

Thus in the riskneutral world, the drifts $\alpha(t,T)$ are completely determined by the volatility surface $\sigma(t,T)$. To create an HJM model, therefore, just follow the following steps:

- Estimate (or otherwise specify) a volatility surface $\sigma(t,T)$.
- Calculate the drifts $\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)^{tr}\, ds$.

- Observe the term structure of forward rates $\{f^*(0,T) : T \geq 0\}$. This involves building a yield curve for all maturities.
- Integrate:
$$f(t,T) = f^*(0,T) + \int_0^t \alpha(u,T)\, du + \int_0^t \sigma(u,T)\, dW_u$$
- Compute bond prices $p(t,T) = e^{-\int_t^T f(t,s)\, ds}$ and the prices of other interest rate derivatives.

8.4.3 Examples and Applications

Example 8.4.5 We consider here the simplest possible HJM model: We have only one source of noise, and put $\sigma(t,T) = \sigma =$ constant for all t, T. By the HJM drift conditions, we see that
$$\alpha(t,T) = \sigma \int_t^T \sigma\, ds = \sigma^2(T-t)$$
under the riskneutral measure. Hence the riskneutral dynamics of forward rates are
$$\begin{aligned} df(t,T) &= \sigma^2(T-t)\, dt + \sigma\, dW_t \\ f(0,T) &= f^*(0,T) \end{aligned}$$
Integrate this to obtain
$$\begin{aligned} f(t,T) &= f^*(0,T) + \sigma^2 t(T - \frac{t}{2}) + \sigma W_t \quad \text{so that} \\ r(t) &= f^*(0,t) + \frac{1}{2}\sigma^2 t^2 + \sigma W_t \end{aligned}$$
and thus the short rate dynamics are given by
$$dr_t = \left[f_T^*(0,t) + \sigma^2 t\right]\, dt + \sigma\, dW_t$$
These short rate dynamics should be familiar: We've obtained the Ho–Lee model fitted to the initial term structure! Note that we didn't have to do the actual fitting — in the HJM framework, fitting is automatic.

$\square$

Thus the Ho–Lee model is (equivalent to) the simplest HJM model.

Example 8.4.6 Can the Hull–White (extended Vasiček) model be recast in the HJM framework?
Indeed it can. The Hull–White model $dr_t = (b(t) - ar_t)\, dt + \sigma\, dW_t$ is an affine term structure model, with bond prices $p(t,T) = e^{A(t,T) - B(t,T)r_t}$. Hence $f(t,T) = -A_T(t,T) + B_T(t,T)r_t$, which means
$$df(t,T) = [\cdot]\, dt + B_T(t,T)\sigma\, dW_t$$
where we haven't bothered to calculate the coefficient of the dt–term (which is, of course, just $\alpha(t,T)$). But for the Hull–White model, it was easy to calculate $B(t,T) = \frac{1}{a}[1 - e^{-a(T-t)}]$, so that $B_T(t,T) = e^{-a(T-t)}$. It follows that
$$df(t,T) = \alpha(t,T)\, dt + \sigma e^{-a(T-t)}\, dW_t$$

Thus $\sigma(t,T) = \sigma e^{-a(T-t)}$ We can now use the HJM drift conditions to calculate $\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)\,ds = \frac{\sigma^2}{a}[e^{-a(T-t)} - e^{-2a(T-t)}]$.

To verify that the above model leads to the Hull–White model, recall that the short rate dynamics can be deduced from the forward rate dynamics as follows:

$$dr_t = [f_T(t,t) + \alpha(t,t)]\,dt + \sigma(t,t)\,dW_t$$

Now $\sigma(t,t) = \sigma$, and $\alpha(t,t) = 0$. Finally,

$$f(t,T) = f(0,T) + \int_0^t \alpha(u,T)\,du + \int_0^t \sigma(u,T)\,dW_u$$

which implies that

$$r(t) = \Theta(t) + \int_0^t \sigma e^{-a(t-u)}\,dW_u$$

for some function $\Theta(t)$, and hence that

$$\begin{aligned} dr(t) &= \Theta'(t)\,dt - \left(a\int_0^t \sigma e^{-a(t-u)}\,dW_u\right)\,dt + \sigma dW_t \\ &= \left[\Theta'(t) - a(r(t) - \Theta(t))\right]\,dt + \sigma\,dW_t \\ &= [b(t) - ar(t)]\,dt + \sigma\,dW_t \end{aligned}$$

Moreover, $b(t) = \Theta'(t) + a\Theta(t)$, and $\Theta(t) = f(0,t) + \int_0^t \alpha(u,t)\,du = f(0,t) + \frac{\sigma^2}{2a^2}(1-e^{-at})^2$. This is exactly the value of $b(t)$ which we obtained for the Hull–White model fitted to the initial term structure.

□

Remarks 8.4.7 The above example suggests a simple mechanism for turning a fitted affine term structure model $dr_t = \mu_r\,dt + \sigma_r\,dW_t$ into an HJM model:

- If $p(t,T) = e^{A(t,T)-B(t,T)r_t}$, solve the (Riccatti) ODE for $B(t,T)$.
- Then the HJM volatility surface is given by $\sigma(t,T) = B_T(t,T)\sigma_r\,dW_t$.
- The HJM drift conditions now specify $\alpha(t,T)$ as well.

□

Example 8.4.8 We consider a model with two sources of noise W_t^1, W_t^2 and a volatility surface

$$\sigma(t,T) = (\sigma_1, \sigma_2 e^{-a(T-t)})$$

where σ_1, σ_2, a are positive constants. The HJM drift conditions dictate that

$$\alpha(t,T) = \sigma_1^2(T-t) + \frac{\sigma_2^2}{a}\left[e^{-a(T-t)} - e^{-2a(T-t)}\right]$$

Integrating the forward rate dynamics, we see

$$\begin{aligned} f(t,T) = f(0,T) + \sigma_1^2 t\left(T - \frac{t}{2}\right) + \frac{\sigma_2^2}{2a^2}\left[2e^{-aT}(1-e^{at}) - e^{-2aT}(1-e^{2at})\right] \\ + \sigma_1 W_t^1 + \sigma_2\int_0^t e^{-a(T-u)}\,dW_u^2 \end{aligned}$$

Thus

$$
\begin{aligned}
r_t &= f(0,t) + \frac{\sigma_1^2 t^2}{2} + \frac{\sigma_2^2}{2a^2}\left[2(e^{-at}-1) - (e^{-2at}-1))\right] \\
&\quad + \sigma_1 W_t^1 + \sigma_2 \int_0^t e^{-a(t-u)}\,dW_u^2 \\
&= \Theta(t) + \sigma_1 W_t^1 + \sigma_2 \int_0^t e^{-a(t-u)}\,dW_u^2
\end{aligned}
$$

Thus the short rate is a Gaussian process, and

$$
\begin{aligned}
dr_t &= \left[\Theta'(t) - a\sigma_2 \int_0^t e^{-a(t-u)}\,dW_u^2\right]\,dt + \sigma_1\,dW_t^1 + \sigma_2\,dW_t^2 \\
&= \left[\Theta'(t) - a(r_t - \Theta(t) - \sigma_1 W_t^1)\right]\,dt + \sigma_1\,dW_t^1 + \sigma_2\,dW_t^2 \\
&= \left[b(t) - ar_t - a\sigma_1 W_t^1\right]\,dt + \sigma_1\,dW_t^1 + \sigma_2\,dW_t^2
\end{aligned}
$$

This is not the form of one of our standard short rate models, because of the explicit presence of W_t^1 in the drift.

□

8.5 Market Models: Preliminaries

The HJM aproach studies the entire term structure of instantaneous forward rates $\{f(t,T) : t \leq T\}$, with considerable success, as we have seen. Nevertheless, forward rates for only a few maturities are available in the market, so the forward rate curve, like the instantaneous short rate, is a purely mathematical entity, a mathematical idealization. Market models, on the other hand, model observable (i.e. market–quoted) rates rather than idealized entities, and thus simple, discrete rates.

The London Interbank Offer Rates (LIBOR), for example, are quoted for different maturities (3–month, 6-month, etc.) and also for different currencies. These LIBOR spot rates imply LIBOR forward rates using an arbitrage argument. New LIBOR quotes are available daily. Swap rates (the fair rates for interest rate swaps) are another example of discrete market–quoted rates. The market model approach to interest rates dates back to Miltersen, Sandmann and Sondermann (1997), Brace, Gatarek and Musiela (1997) and Jamshidian (also 1997). Several other approaches now exist, due to Hunt and Kennedy, and Musiela and Rutkowski, amongst others. It remains one of the most intensively researched areas of financial mathematics.

8.5.1 Black's Models

Black's model has long been the industry–standard model used by traders to price a variety of European–style options, including interest rate options, such as caps,floors, and swaptions. It is essentially a minor variation on the Black–Scholes formula, as we shall shortly see. Nevertheless, the suitability and adequacy of Black's model has often been questioned by academics, particularly in the arena of interest rate options.

Consider a European call option C with strike K and maturity T on some market variable X. X need not be a traded instrument — it could also be a market–quoted interest rate, for example. The main assumption is that X_T is lognormally distributed *in the riskneutral*

world.. Thus we make no assumptions on the distribution of the process $(X_t)_t$ in general, but just on the value of X at the expiry of the option. We further define the "volatility" of X_T to be a non–negative number σ satisfying

$$\text{variance of } \ln X_T = \sigma^2 T$$

Let $\mathbb{Q}$ be a riskneutral measure. Then the $t = 0$–value of the call is

$$C_0 = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}(X_T - K^+)\right]$$

Black uses two approximations to determine the value of C_0:

- Approximate

  $$\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}(X_T - K)^+\right] \approx P(0,T)\mathbb{E}_{\mathbb{Q}}[(X_T - K)^+]$$

  i.e. discount outside the expectation operator.

- Now because X_T is lognormal under $\mathbb{Q}$, we know that

  $$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[(X_T - K)^+] &= \mathbb{E}[X_T]N(d_1) - KN(d_2) \qquad \text{where} \\ d_1 &= \frac{\ln \frac{\mathbb{E}_{\mathbb{Q}}[X_T]}{K} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

  Approximate

  $$\mathbb{E}_{\mathbb{Q}}[X_T] = \text{forward price/rate of } X = F_0$$

  i.e. approximate the expectation by the forward price/rate.
  Since the forward price of X at time T for time T is just itself (i.e. $F_T = X_T$), this can be interpreted as saying that the forward rate process has zero drift, i.e. is a $\mathbb{Q}$–martingale.

Thus, using these two approximations, we obtain Black's model for a call on X:

$$\begin{aligned} C_0 &= P(0,T)[F_0 N(d_1) - KN(d_2)] \qquad \text{where} \\ d_1 &= \frac{\ln \frac{F_0}{K} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

A similar formula is obtained for puts, using put–call parity.

If payments are based on a variable X_T, but only received at some later date T^*, then discounting must be done from time T^* rather than from time T. Black's model then generalizes to give call prices

$$\begin{aligned} C_0 &= P(0,T^*)[F_0 N(d_1) - KN(d_2)] \qquad \text{where} \\ d_1 &= \frac{\ln \frac{F_0}{K} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

where F_0 is still the T–forward value of X at time $t = 0$. The appropriate generalized Black formula for put options follows once again by put–call parity.

Now it ought to be clear Black's model has several flaws. Firstly, it cannot be appropriate to use the first approximation when X_T depends on interest rates, as it amounts to saying that

$$\mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}(X_T-K)^+\right] = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}\right]\mathbb{E}_{\mathbb{Q}}\left[(X_T-K)^+\right]$$

which is close to asserting that r and X_T are independent. That's a dangerous assumption if X happens to be an interest rate derivative! There is no justification for the second approximation either. The expected value of X_T under the riskneutral measure is its *futures price*, whereas the forward price is the expected value of X_T under the T–forward riskneutral measure. These measures are not the same if interest rates are stochastic.

In spite of these flaws, Black's model remains heavily used — the industry standard. The method can be justified, provided that the relevant variable is taken to be lognormal under a different measure, associated with a different numéraire. We shall give several examples of this below. Review material on changes of measure and numéraire may be found in the next subsection.

Example 8.5.1 Bond Options: Lognormal prices
We consider a call C with strike K and maturity T on a coupon bearing bond B. We assume that the bond price at time T is lognormally distributed (under the riskneutral measure), and that $\ln B_T$ has variance $\sigma^2 T$. This "volatility" σ is obtained from historical data (or implied by other market variables).[1]

The T–forward bond price F_0 is simply the fair price which sets the value of a forward contract on B equal to zero. A simple arbitrage argument shows

$$F_0 = \frac{B_0 - D}{P(0,T)}$$

where B_0 is the current value of the bond, $P(0,T)$ is the discount bond maturing at time T, and D is the present value of all coupons (dividends) paid out during the life of the option. Thus Black's model determines

$$\begin{aligned} C_0 &= (B_0 - D)N(d_1) - KP(0,T)N(d_2) \qquad \text{where} \\ d_1 &= \frac{\ln \frac{B_0 - D}{KP(0,T)} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

The above call price is an *approximation* under the assumption that B_T is lognormal under $\mathbb{Q}$, but *exact* if we assume lognormality of B_T under the T–forward riskneutral measure $\mathbb{Q}^T$.

Of course, bond put options can be evaluated by put–call parity.

□

[1] In practice, yield volatilities are often obtained. If σ_y is the volatility of the yield (i.e. if $\sigma_y^2 T$ is the standard deviation of the logarithm of the forward yield $\ln y_T$), then (with $D^* =$ duration) we have $\frac{\Delta B}{B_0} \approx -D^*\Delta y = -D^* y_0 \frac{\Delta y}{y_0}$, i.e. $\Delta(\ln B) \approx -D^* y_0 \Delta(\ln y)$. Thus the variance of $\ln B$ is approximately $(D^* y_0)^2\times$ the variance of $\ln y$, i.e. $\sigma_B \approx D^* y_0 \sigma_y$.

Example 8.5.2 Caps: Lognormal LIBOR Rates
An interest rate cap is an option–like contract which protects the holder against a floating interest rate moving too high. Each cap is a portfolio of caplets, each for a certain future time interval. A caplet is essentially a call option on the floating rate, given a certain cap rate as strike, based on a given notional amount. Consider, for example, a five–year cap, on a notional amount A, with cap rate R and semiannual resets based on 6–month LIBOR. This is a portfolio of 10 caplets. The reset dates $T_0 = 0, T_1 = 0.5, T_2 = 1, \dots T_{10} = 5$ are referred to as the tenor structure of the cap. The n^{th} caplet protects the holder against 6–month LIBOR rising above R over the period $[T_{n-1}, T_n]$. It is a call option with strike R on the 6–month spot LIBOR $L(T_{n-1})$ at time T_{n-1}, and will have the following payoff at time T_n:

$$\text{Payoff of } n^{\text{th}} \text{ cap} = A\delta_n (L(T_{n-1}) - R)^+ \qquad \text{where } \delta_n = T_n - T_{n-1}$$

(This is a *payment–in–arrears* cap. The first caplet is generally excluded from the cap, because there is no uncertainty about the spot LIBOR $L(T_0)$.)

To price the n^{th} caplet using Black's model, we assume that the future spot LIBOR $L(T_{n-1})$ is lognormally distributed, with volatility σ_{n-1}. The $t = 0$–forward LIBOR rate (i.e. the F_0 of Black's model) for the period $[T_{n-1}, T_n]$ is given by

$$L(0, T_{n-1}) = \frac{P(0, T_{n-1}) - P(0, T_n)}{\delta_n P(0, T_n)}$$

(In this notation, the future spot rate, $L(T_{n-1})$, is just $L(T_{n-1}, T_{n-1})$.)
Hence the $t - 0$–value of the n^{th} caplet is

$$\begin{aligned}
C_n(0) &= A\delta_n P(0, T_n) \left[L(0, T_{n-1}) N(d_{1,n-1}) - RN(d_{2,n-1}) \right] \\
d_{1,n-1} &= \frac{\ln \frac{L(0,T_{n-1})}{R} + \frac{1}{2}\sigma_{n-1}^2 T_{n-1}}{\sigma \sqrt{T_{n-1}}} \\
d_{2,n-1} &= d_{1,n-1} - \sigma_{n-1}\sqrt{T_{n-1}}
\end{aligned}$$

The price of the cap is therefore the *sum of the prices of the caplets* (though, as we have mentioned, the first cap is often excluded, i.e. $C_1(0)$ is set to zero).

The above price for a cap is an *approximation*, assuming that each future LIBOR spot rate $L(T_n)$ is lognormal under the riskneutral measure $\mathbb{Q}$. The formula for each caplet is *exact*, however, if it is assumed that $L(T_n)$ is lognormal under the T_{n+1}–forward measure. For then indeed

$$\frac{C_n(0)}{P(0, T_n)} = \mathbb{E}_{\mathbb{Q}^{T_n}} \left[\frac{A\delta_n [L(T_{n-1}) - R]^+]}{P(T_n, T_n)} \right]$$

which justifies the first approximation used in Black's model (i.e. discounting outside the expectation). Moreover, the second approximation is exact, i.e. the *forward* LIBOR rate $L(0, T_{n-1})$ is exactly equal to the expected value of the spot rate, but under the forward riskneutral measure: $L(0, T_{n-1}) = \mathbb{E}_{\mathbb{Q}^{T_n}}[L(T_{n-1})]$. To see this, note that a long forward rate agreement F, initiated at time $t = 0$ for period $[T_{n-1}, T_n]$, will have initial value $F_0 = 0$, and terminal value $F_{T_n} = \delta_n[L(T_{n-1}) - L(0, T_{n-1})]$. Hence

$$0 = \frac{F_0}{P(0, T_{n-1})} = \mathbb{E}_{\mathbb{Q}^{T_n}} \left[\frac{F_{T_n}}{P(T_n, T_n)} \right]$$

which yields the required result (because $L(0, T_{n-1})$ is a known constant).

So in order for the Black price of a cap to be accurate, we must simultaneously assume that each $L(T_n)$ is lognormal under $\mathbb{Q}^{T_{n+1}}$. This seems difficult to justify theoretically. One of the achievements of LIBOR market models is that they provide a framework under which these assumptions all *do* hold simultaneously, thus showing that the use of Black's model does not lead automatically to arbitrage opportunities.

□

Example 8.5.3 Caps: Lognormal Bond Prices
A cap can be decomposed into a portfolio of puts on zero coupon bonds. To be precise, the n^{th} caplet (from the previous example) has

$$\text{Payoff} = A\delta_n[L(T_{n-1}) - R]^+ \quad \text{at time } T_n$$

Since $L(T_{n-1})$ is known at time T_{n-1} this is equivalent to a time–T_{n-1} payoff of

$$\begin{aligned}\frac{A\delta_n[L(T_{n-1}) - R]^+}{1 + \delta_n L(T_{n-1})} &= A\left[1 - (1 + \delta_n R)P(T_{n-1}, T_n)\right]^+ \\ &= A(1 + \delta_n R)\left[\frac{1}{1 + \delta_n R} - P(T_{n-1}, T_n)\right]^+\end{aligned}$$

This last line is easily seen to be the time–T_{n-1} payoff of a portfolio of $A(1+\delta_n R)$–many put options with strike $\frac{1}{1+\delta_n R}$ and expiry T_{n-1} on underlying security $P(t, T_n)$. If at time T_{n-1} the caplet has the same payoff as a portfolio of puts on $P(t, T_n)$, then, by the Law of One Price, the value of the caplet must have the same value as the portfolio of puts at any earlier time as well.

Thus the $t = 0$–value of the n^{th} caplet is

$$\begin{aligned}C_n(0) = A(1 + \delta_n R) \times \text{ value of put option on } P(t, T_n) \text{ with strike } \frac{1}{1 + \delta_n R} \\ \text{and expiry } T_{n-1}\end{aligned}$$

This can be evaluated using the method of the first example of this subsection.

□

Example 8.5.4 Swaptions: Lognormal Swap Rates
Suppose we initiate, at time t, a pay–fixed interest rate swap starting at time $T \geq t$, with tenor structure $T = T_0 < T_1 < \cdots < T_N$ on a notional amount A. This is known as a *forward swap* or *deferred swap*. Let $\delta_n = T_n - T_{n-1}$, and recall that at T_n pay–fixed receives

$$A\delta_n(L(T_{n-1}) - S_{t,T}) \qquad n = 1, \ldots, N$$

where $S_{t,T}$ is the T–forward swap rate at time t, and $L(T_{n-1})$ is the spot LIBOR rate at time T_{n-1} for the period $[T_{n-1}, T_n]$. Further recall that $S_{t,T}$ is the rate which sets the initial (i.e. time t) value of the forward swap equal to zero.

The interest payments on a pay–fixed swap are equivalent to the payments of a portfolio consisting of short a coupon bond with coupon rate $S_{t,T}$, and long a floating rate note. The

bond and the FRN both come into existence at time T. The current value of such a forward starting bond bond is

$$A[\sum_{n=1}^{N} \delta_n S_{t,T} P(t,T_n) + P(t,T_N)]$$

The floating rate note will trade at par at time T, i.e. we need to set aside $AP(t,T)$ at time t to purchase the FRN at time T. Hence the forward swap rate satisfies

$$-A[\sum_{n=1}^{N} \delta_n P(t,T_n) S_{t,T} + P(t,T_N)] + AP(t,T) = 0$$

(where the coupon bond and FRN have the same payment dates as the swap, and the same notional) and thus

$$S_{t,T} = \frac{P(t,T) - P(t,T_N)}{\sum_{n=1}^{N} \delta_n P(t,T_n)}$$

If $t = T$, then $S_{t,t}$ is just the ordinary spot swap rate at time t.

A swaption C is the right to enter into a pay–fixed swap at some future date T at a strike rate R. If the tenor structure is $T = T_0 < T_1 < T_2 < \cdots < T_N$, then the swaption gives the holder the right (but not the obligation) to receive at each of the dates $T_1, \ldots, T_N$ an amount

$$A\delta_n(L(T_{n-1}) - R)$$

If a pay–fixed swap were to be entered at time T at the spot swap rate, then payments would be

$$A\delta_n(L(T_{n-1}) - S_{T,T})$$

and thus the swaption would be exercised only if $R < S_{T,T}$. The swaption thus gives rise to a series of payments

$$A\delta_n(S_{T,T} - R)^+$$

at times T_n. Each payment is equivalent to the payoff of $A\delta_n$–many calls with strike R and maturity T on underlying $S_{T,T}$. Using the generalized version of Black's model, i.e. assuming that $S_{T,T}$ is lognormal under the riskneutral measure and making the appropriate approximations, the $t = 0$–value of each such payment is

$$A\delta_n P(0,T_n)\,[S_{0,T} N(d_1) - RN(d_2)]$$

where $d_1 = \frac{\ln \frac{S_{0,T}}{R} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}$, $d_2 = d_1 - \sigma\sqrt{T}$, and σ is the volatility of the future spot swap rate $S_{T,T}$. Hence the value of the swaption is

$$\begin{aligned}
C_0 &= \sum_{n=1}^{N} A\delta_n P(0,T_n)[S_{0,T} N(d_1) - RN(d_2)] \qquad \text{where} \\
d_1 &= \frac{\ln \frac{S_{0,T}}{R} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\
d_2 &= d_1 - \sigma\sqrt{T}
\end{aligned}$$

and

$$S_{0,T} = \frac{P(t,T) - P(t,T_N)}{\sum_{n=1}^{N} \delta_n P(0,T_n)}$$

We saw that we can make the Black formula for caps exact, provided we work with the appropriate numéraires, under the appropriate equivalent martingale measures. Can we make Black's formula for swaptions exact? Yes, indeed. Note that the numerator in the expression for $S_{0,T}$ is equivalent to a portfolio of zero coupon bonds, i.e.

$$\sum_{n=1}^{N} \delta_n P(0, T_n)$$

corresponds to a stream of cashflows of size δ_n at time T_n. If, as is often the case, all the δ_n are of the same size, then this portfolio is just an annuity. Now we may think of the portfolio as a traded asset, call it X, and use it as numéraire.

The first of the Black approximations is exact under the measure $\mathbb{Q}_X$: The time–T value of all the payoffs of the swaption is

$$C_T = \sum_{n=1}^{N} A\delta_n P(T, T_n)[S_{T,T} - R]^+ = AX_T[S_{T,T} - R]^+$$

Hence

$$\frac{C_0}{X_0} = \mathbb{E}_{\mathbb{Q}_X}\left[\frac{C_T}{X_T}\right] = \mathbb{E}_{\mathbb{Q}_X}[A(S_{T,T} - R)^+]$$

so that

$$C_0 = \sum_{n=1}^{N} A\delta_n P(0, T_n)\mathbb{E}_{\mathbb{Q}_X}[(S_{T,T} - R)^+]$$

i.e. we discount outside the expectation.
As for the second approximation, we need to show that the forward swap rate $S_{0,T}$ (which can now be seen to equal $\frac{P(0,T)-P(0,T_N)}{X_0}$) is just the expected value of the future spot swap rate $S_{T,T}$ under the EMM $\mathbb{Q}_X$, i.e. that $\mathbb{E}_{\mathbb{Q}_X}[S_{T,T}] = S_{0,T}$. To see this, consider a pay–fixed forward swap F initiated at $t = 0$ to start at time T, with interest payment dates $T_1, \ldots, T_N$. The $t = 0$–value of the contract is $F_0 = 0$, whereas at time T the value is $F_T = \sum_{n=1}^{N} A\delta_n P(T, T_n)[S_{T,T} - S_{0,T}] = AX_T[S_{T,T} - S_{0,T}]$. The desired result now follows immediately from the fact that

$$0 = \frac{F_0}{X_0} = \mathbb{E}_{\mathbb{Q}_X}\left[\frac{F_T}{X_T}\right]$$

Hence Black's formula is exact, provided we assume that swap rates are lognormally distributed under the EMM associated with the annuity process $X_t = \sum_{n=1}^{N} \delta_n P(t, T_n)$.

□

It's pretty amazing that the Black formula for various derivatives (published in 1976) can in many cases be made exact using the change of numéraire technique (discovered in the early 1990's). In particular, both the Black formula for caps and that for swaptions are exact if we assume that LIBOR rates are lognormal under the appropriate forward riskneutral measures, and that swap rates are lognormal under the "annuity" measure.

8.5.2 Review of Changes of Measure and Numéraire; LIBOR Rates

Fix a horizon $T^* > 0$ and suppose that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t, (S_t^i)_{i,t})$ is a market model, where the filtration $(\mathcal{F}_t)_t$ is generated by a standard (multi–dimensional) $\mathbb{P}$–Brownian motion $(W_t)_t$, augmented to satisfy the usual conditions. Let $\mathbb{Q}$ be the riskneutral measure, i.e. a measure which has the property that all asset price processes S_t^i are martingales when denominated in units of the money market account A_t. We briefly recall some facts about how Girsanov's Theorem is used to change the measure (e.g. to construct $\mathbb{Q}$ from $\mathbb{P}$):

- Assume that the asset dynamics are given by

$$\frac{dS_t^i}{S_t^i} = \mu^i(t, S_t)\,dt + \sigma^i(t, S_t)\,dW_t \qquad \frac{dA_t}{A_t} = r_t A_t\,dt$$

  with suitable initial conditions.
  Recall that the market price of risk $\lambda_t^{\mathbb{P}}$ is a vector satisfying

$$\sigma_t^i \cdot \lambda_t^{\mathbb{P}} = \mu_t^i - r_t$$

  (This looks like it depends on the asset S^i, but we know from previously developed theory that, for a model to be arbitrage–free, all assets must have the same market price of risk. Hence we've suppressed an index i.)

- Let $u(t, \omega)$ be a predictable process, to be used as a kernel for a Girsanov transformation.

- Define a new measure $\tilde{\mathbb{P}}$ by

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \mathcal{E}_{T^*}\left(\int_0^{\cdot} u_t\,dW_t\right) = e^{\int_0^T u_t\,dW_t - \frac{1}{2}\int_0^T ||u_t||^2\,dt}$$

- Girsanov's Theorem states that

$$\tilde{W}_t = W_t - \int_0^t u_s\,ds$$

  is a $\tilde{\mathbb{P}}$–Brownian motion.

- Thus the new asset dynamics are, under $\tilde{\mathbb{P}}$,given by

$$\frac{dS_t^i}{S_t^i} = (\mu_t^i + \sigma_t^i u_t)\,dt + \sigma_t^i\,d\tilde{W}_t \qquad \frac{dA_t}{A_t} = r_t A_t\,dt$$

  It follows that the market price of risk under $\tilde{\mathbb{P}}$ must satisfy the relation

$$\lambda_t^{\tilde{\mathbb{P}}} \sigma_t^i = \mu_t^i + \sigma_t^i u_t - r_t = (\lambda_t^{\mathbb{P}} + u_t)\sigma_t^i$$

  and thus

$$\lambda_t^{\tilde{\mathbb{P}}} = \lambda_t^{\mathbb{P}} + u_t$$

- Hence a Girsanov transformation adds the Girsanov kernel to the market price of risk. It adds volatility × kernel to the drift.

- To obtain a riskneutral measure $\mathbb{Q}$, the new market price of risk $\lambda_t^{\mathbb{Q}}$ must be zero, and thus we must have $u_t = -\lambda_t^{\mathbb{P}}$. This is in agreement with what we found earlier. In that case, the drift becomes $\mu^i - \sigma_t^i \lambda_t^{\mathbb{P}} = r_t$, which we already know very well.

- To change from the riskneutral measure $\mathbb{Q}$ to an equivalent martingale measure $\mathbb{Q}_X$ for numéraire X, we proceed as follows: Start in the riskneutral world, where $\frac{dS^t}{S_t} = r\ dt + \sigma_S\ dW_t^{\mathbb{Q}}$, and $\frac{dX_t}{X_t} = r\ dt + \sigma_X\ dW_t^{\mathbb{Q}}$. Under $\mathbb{Q}_X$, the ratios $\hat{S}_t = \frac{S_t}{X_t}$ are martingales. Now under $\mathbb{Q}$, the ratios have dynamics

$$\begin{aligned}\frac{d\hat{S}_t}{\hat{S}_t} &= -\sigma_X(\sigma_S - \sigma_X)\ dt + (\sigma_S - \sigma_X)\ dW_t^{\mathbb{Q}} \\ &= -\sigma_X \hat{\sigma}\ dt + \hat{\sigma}\ dW_t^{\mathbb{Q}}\end{aligned}$$

(where $\hat{\sigma} = \sigma_S - \sigma_X$). To make the drift equal to zero (i.e. to make $\hat{S}_t$ into a martingale), we need to to add $\hat{\sigma} \times \sigma_X =$ volatility $\times \sigma_X$, i.e. we need to use a Girsanov transformation with kernel σ_X. Thus

$$\frac{d\mathbb{Q}_X}{d\mathbb{Q}} = \mathcal{E}_T(\int_0^{\cdot} \sigma_X\ dW_t^{\mathbb{Q}})$$

- Hence we need to add σ_X to the riskneutral market price of risk to obtain the market price of risk under $\mathbb{Q}_X$. Since the riskneutral market price of risk is zero, the market price of risk under $\mathbb{Q}_X$ is just the volatility of the numéraire X.

- Numéraire–denominated asset price dynamics under the associated equivalent martingale measure are therefore just

$$\frac{d\hat{S}_t}{S_t} = (\sigma_S - \sigma_X)\ dW_t^{\mathbb{Q}_X}$$

- If the numéraire is the T–bond $P(t,T)$, the associated EMM is called the T–forward riskneutral measure, and denoted by $\mathbb{Q}^T$. If bond price dynamics are

$$\frac{dP(t,S)}{P(t,S)} = \mu_S(t)\ dt + \sigma_S(t)\ dW_t^{\mathbb{P}}$$

under the "real–world" measure $\mathbb{P}$, then the numéraire denominated dynamics are given by

$$\frac{d\hat{P}(t,S)}{\hat{P}(t,S)} = (\sigma_S - \sigma_T)\ dW_t^T$$

where $\hat{P}(t,S) = \frac{P(t,S)}{P(t,T)}$ and W_t^T is a $\mathbb{Q}^T$–Brownian motion.

- Given future times $T < S$, the market price of risk under $\mathbb{Q}^S$ is just σ_S, whereas the market price of risk under $\mathbb{Q}^T$ is σ_T. To move from $\mathbb{Q}^S$–world to $\mathbb{Q}^T$–world, we must change the market price of risk from σ_S to σ_T, i.e. we need to add $\sigma_T - \sigma_S$ to the market price of risk under $\mathbb{Q}^S$. Hence the change from $\mathbb{Q}^S$–world to $\mathbb{Q}^T$–world is effected by a Girsanov transformation with kernel $\sigma_T - \sigma_S$, i.e.

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}^S} = \mathcal{E}_T\left(\int_0^{\cdot} \sigma_T - \sigma_S\ dW_t^{\mathbb{Q}^S}\right)$$

We can also verify this directly. Recall that the Radon–Nikodym process $\xi_t = \mathbb{E}_{\mathbb{Q}^S}[\frac{d\mathbb{Q}^T}{d\mathbb{Q}^S}|\mathcal{F}_t]$ for a change of numéraire is given by a ratio of asset ratios:

$$\xi_t = \frac{P(t,T)/P(t,S)}{P(0,T)/P(0,S)}$$

and thus

$$d\xi_t = \xi_t[\sigma_T(t) - \sigma_S(t)]\, dW_t^{\mathbb{Q}^S}$$

The solution of this SDE, together with the initial condition $\xi_0 = 1$, is just $\xi_t = \mathcal{E}_t\left(\int_0^\cdot \sigma_T(u) - \sigma_S(u)\, dW_u^{\mathbb{Q}^S}\right)$.

- Finally, note that the asset ratio process $\check{P}(t,T) = \frac{P(t,T)}{P(t,S)}$ satisfies the same SDE as does ξ_t

$$\frac{d\check{P}(t,T)}{\check{P}(t,T)} = [\sigma_T(t) - \sigma_S(t)]\, dW_t^{\mathbb{Q}^S}$$

although their initial conditions may differ. Hence ξ_t and $\check{P}(t,T)$ differ by a constant factor, i.e.

$$\mathbb{E}_{\mathbb{Q}^S}\left[\frac{d\mathbb{Q}^T}{d\mathbb{Q}^S}|\mathcal{F}_t\right] = \xi_t = c\check{P}(t,T) = \frac{cP(t,T)}{P(t,S)}$$

Let $T^* > 0$ be a a horizon for our bond market model. The time–t forward LIBOR rate for the future interval $[T, T+\delta]$ (where $T \leq T^* - \delta$) is defined by

$$1 + \delta L(t,T) = \frac{P(t,T)}{P(t,T+\delta)} \quad \text{i.e.} \quad L(t,T) = \frac{P(t,T) - P(t,T+\delta)}{\delta P(t,T+\delta)}$$

We saw earlier that $L(t,T)$ is the interest rate for the period $[T, T+\delta]$ that can be locked in at time t (by a judicious investment in a portfolio of T– and $T+\delta$–bonds with zero initial cost).

Alternatively, the forward LIBOR rate $L(t,T)$ can be regarded as the swap rate for a *single–period swap settled in arrears*. For suppose that we have a single–period interest rate swap, contracted at time t, for the period $[T, T+\delta]$, to be settled at time $T+\delta$. Thus, at time $T+\delta$, the pay–fixed side pays δR, and the receive–fixed party pays $P^{-1}(T, T+\delta) - 1$, where R is the fair swap rate, and $P^{-1}(T, T+\delta) = 1 + \delta S$, $S = L(T,T)$ the spot rate at time T for period $T, T+\delta]$. Equivalently, by adding 1 to both payments, pay–fixed pays Y^{fx} and receive–fixed pays Y^{fl}, where

$$Y^{fx} = 1 + \delta R \qquad Y^{fl} = P^{-1}(T, T+\delta)$$

We can regard Y^{fx} and Y^{fl} as contingent claims which are paid out at time $T+\delta$. It is clear that the time t–value of Y^{fx} is just

$$Y^{fx} = P(t, T+\delta)[1 + \delta R]$$

The time–t value of Y^{fl} is obtained as follows: If, at time T, we invest \$1.00 in $T+\delta$–bonds, the payoff at time $T+\delta$ will be $P^{-1}(T, T+\delta)$. To obtain the required \$1.00, we must invest in one T–bond at time $t \leq T$. Hence

$$Y_t^{fl} = P(t,T)$$

The swap rate at time t is the rate R for which $Y^{fx} = Y^{fl}$, and thus $R = \frac{P(t,T)-P(t,T+\delta)}{\delta P(t,T+\delta)} = L(t,T)$.

Define

$$P(t,T,S) = \frac{P(t,T)}{P(t,S)} = 1 + \delta L(t,T) \qquad \text{for } t \leq T \leq S \text{ and } \delta = S - T$$

Then $P(t,T,S)$ is a $\mathbb{Q}^S$–martingale. In particular, the LIBOR forward rate $L(t,T)$ is a $\mathbb{Q}^{T+\delta}$–martingale. Thus the LIBOR forward rate $L(t,T)$ is simply the expected value of the LIBOR spot rate $L(T,T)$ at time T, where the expectation is taken under the $\mathbb{Q}^{T+\delta}$–measure.

8.6 Lognormal Forward LIBOR Market Models

We start with a pre–specified sequence of times

$$0 = T_0 < T_1 < T_2 < \cdots < T_N = T^*$$

These times, typically settlement– or reset dates, are collectively known as the *tenor structure*. We also define $\delta_j = T_j - T_{j-1}$ for $j = 1, \ldots, N$. Then the forward LIBOR rate satisfy

$$1 + \delta_{j+1} L(t,T_j) = \frac{P(t,T_j)}{P(t,T_{j+1})} = P(t,T_j,T_{j+1})$$

We assume that the bond market satisfies a strong form of the no–arbitrage condition, i.e. we assume that there exists a riskneutral measure $\mathbb{Q}$ simultaneously for all discount bonds $P(t,T)$. We denote, for each $P(t,T)$, its associated forward riskneutral measure by $\mathbb{Q}^T$. W_t and W_t^T will denote, respectively, $\mathbb{Q}$– and $\mathbb{Q}^T$–Brownian motions.

Let $S(t,T)$ be the volatility of the T–bond $P(t,T)$ at time t. From the previous subsection, we know the following:

- $\mathbb{Q}^{T_j}$ is obtained from $\mathbb{Q}^{T_{j+1}}$ via a Girsanov transformation with kernel $S(t,T_j) - S(t,T_{j+1})$, i.e.
$$\frac{d\mathbb{P}^{T_j}}{d\mathbb{P}^{T_{j+1}}} = \mathcal{E}_{T_j}\left(\int_0^{\cdot} S(t,T_j) - S(t,T_{j+1})\, dW_t^{T_{j+1}}\right)$$
- Each asset ratio $P(t,T_j,T_{j+1}) = \frac{P(t,T_j)}{P(t,T_{j+1})}$ is a $\mathbb{Q}^{T_{j+1}}$–martingale.
- Each forward LIBOR rate $L(t,T_j)$ is a $\mathbb{Q}^{T_{j+1}}$–martingale.
- The $\mathbb{Q}^{T_{j+1}}$–dynamics of the asset ratio $P(t,T_j,T_{j+1})$ are
$$\frac{dP(t,T_j,T_{j+1})}{P(t,T_j,T_{j+1})} = (S(t,T_j) - S(t,T_{j+1}))\, dW_t^{T_{j+1}}$$
- There is a constant c such that the Radon–Nikodym process and the asset ratio process are related
$$\mathbb{E}_{\mathbb{Q}^{T_{j+1}}}\left[\frac{d\mathbb{Q}^{T_j}}{d\mathbb{Q}^{T_{j+1}}}|\mathcal{F}_t\right] = cP(t,T_j,T_{j+1}) = c(1 + \delta_{j+1} L(t,T_j))$$

Note that, assuming that the forward LIBOR rate processes $L(t,T)$ are strictly positive, we have the following dynamics:

$$dL(t,T_j) = L(t,T_j)\lambda(t,T_j)\, dW_t^{T_{j+1}}$$

This follows from the Martingale Representation Theorem: $L(t,T_j)$ is a $\mathbb{Q}^{T_{j+1}}$–martingale, and thus we must have $dL(t,T_j) = h_t\, dW_t^{T_{j+1}}$. Since $L(t,T_j)$ is strictly positive, we may define $\lambda(t,T_j) = frach_t L(t,T_j)$ to obtain $dL(t,T_j) = L(t,T_j)\lambda(t,T_j)\, dW_t^{T_{j+1}}$.

Now $P(t,T_j,T_{j+1}) = 1+\delta_{j+1}L(t,T_j)$, so that $dP(t,T_j,T_{j+1}) = \delta_{j+1}\, dL(t,T_j) = \delta_{j+1}L(t,T_j)\lambda(t,T_j)\, dW_t^{T_{j+1}}$. We also found that $\frac{dP(t,T_j,T_{j+1})}{P(t,T_j,T_{j+1})} = (S(t,T_j) - S(t,T_{j+1}))\, dW_t^{T_{j+1}}$, and equating these expressions yields

$$\frac{\delta_{j+1}L(t,T_j)}{1+\delta_{j+1}L(t,T_j)}\lambda(t,T_j) = S(t,T_j) - S(t,T_{j+1})$$

This expression will play an important role in the inductive construction of lognormal models of forward LIBOR rates.

Since the move from $\mathbb{Q}^{T_{j+1}}$–world to $\mathbb{Q}^{T_j}$–world is accomplished by a Girsanov transformation with kernel $S(t,T_j) - S(t,T_{j+1}) = \frac{\delta_{j+1}L(t,T_j)\lambda(t,T_j)}{1+\delta_{j+1}L(t,T_j)}$, the dynamics of $L(t,T_j)$ under $\mathbb{Q}^{T_j}$ are given by

$$dL(t,T_j) = L(t,T_j)\left[\frac{\delta_{j+1}||\lambda(t,T_j)||^2}{1+\delta_{j+1}L(t,T_j)}\, dt + \lambda(t,T_j)\, dW_t^{T_j}\right]$$

because volatility $\times$ kernel must be added to the $\mathbb{Q}^{T_{j+1}}$–drift of $L(t,T_j)$, while leaving the volatility unchanged (and the drift is zero, while the volatility is $\lambda(t,T_j)$).

8.6.1 The Brace–Gatarek–Musiela Approach to Forward LIBOR

In most markets, caps and floors form the largest component of an average swap derivatives book.... Market practice is to price the option assuming that the underlying forward rate process is lognormally distributed with zero drift. Consequently, the option price is given by the Black futures formula, discounted from the settlement data.

In an arbitrage–free setting, forward rates over consecutive intervals are all related to one another, and cannot all be lognormal under one arbitrage–free measure. That is probably what led the academic community to a degree of skepticism toward the market practice of pricing caps...

The aim of this paper is to show that market practice can be made consistent with an arbitrage–free term structure model... This is possible because each rate is lognormal under the forward (to the settlement date) arbitrage–free measure rather than under one (spot) arbitrage–free measure. Lognormality under the appropriate forward and not spot arbitrage–free measure is needed to justify the Black futures formula with discount for caplet pricing.

— Brace, Gatarek, Musiela [1997]

The BGM–model starts from a family $P(t,T)$ of discount bond prices up to some horizon maturity T^*. We assume that each forward rate is over a period of length δ (the same for all

rates). The bond price processes also give us the bond ratio processes (i.e. forward prices) $P(t,T,S) = \frac{P(t,T)}{P(t,S)}$. The forward LIBOR rates $L(t,T)$ are thus defined by

$$1 + \delta L(t,T) = P(t,T,T+\delta) \qquad \text{for } T \leq T^* - \delta$$

BGM put their model inside the HJM framework, i.e. they assume that a term structure of instantaneous forward rates for all maturities (less than the horizon date T^*) is available. In contrast, the Musiela–Rutkowski and Jamshidian approaches require forward rates only for a discrete set of tenor dates, as we shall see. Now recall that if, in an HJM model, the riskneutral dynamics of the instantaneous forward rate $f(t,T)$ is given by

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t \quad \text{where } \alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,du$$

(using the HJM drift condition), then the riskneutral bond price dynamics are given by

$$\frac{dP(t,T)}{P(t,T)} = r_t\,dt + S(t,T)\,dW_t \quad \text{where } S(t,T) = -\int_t^T \sigma(t,u)\,du$$

Further recall that earlier we obtained

$$\frac{\delta L(t,T)}{1+\delta L(t,T)}\lambda(t,T) = S(t,T) - S(t,T+\delta)$$

(which also follows if we apply Itô's formula to the identity $1 + \delta L(t,T) = e^{\int_T^{T+\delta} f(t,u)\,du}$ and compare the dW_t–terms). The main problem is this:

> How can we specify bond volatilities $S(t,T)$ (or equivalently, the instantaneous forward rate volatilities $\sigma(t,T) = -\frac{\partial S(t,T)}{\partial T}$) so that the resulting *discrete simple* forward LIBOR rates will have the desired deterministic volatity structure?

We have already seen that $L(t,T)$ is a non–negative $\mathbb{Q}^{T+\delta}$–martingale. For $1+\delta L(t,T) = P(t,T,T+\delta)$, and so $dL(t,T) = \delta^{-1} dP(t,T,T+\delta)$. But $P(t,T,T+\delta)$ is a $\mathbb{Q}^{T+\delta}$–martingale (by definition of $\mathbb{Q}^{T+\delta}$), with dynamics $\frac{dP(t,T,T+\delta)}{P(t,T,T+\delta)} = [S(t,T) - S(t,T+\delta)]\,dW_t^{T+\delta}$. It therefore follows that

$$\begin{aligned} dL(t,T) &= \delta^{-1} P(t,T,T+\delta)[S(t,T) - S(t,T+\delta)]\,dW_t^{T+\delta} \\ &= L(t,T)\left(\frac{[1+\delta L(t,T)][S(t,T) - S(t,T+\delta)]}{\delta L(t,T)}\right)\,dW_t^{T+\delta} \end{aligned}$$

i.e.

$$dL(t,T) = L(t,T)\lambda(t,T)\,dW_t^{T+\delta}$$
$$\text{where } \lambda(t,T) = \frac{[1+\delta L(t,T)][S(t,T) - S(t,T+\delta)]}{\delta L(t,T)}$$

We are therefore able to derive the forward LIBOR dynamics directly from the bond price volatilities (or, equivalently, the instantaneous forward rate volatilities). Since the forward

riskneutral measure $\mathbb{Q}^{T+\delta}$ is obtained from the (spot) riskneutral measure $\mathbb{Q}$ by a Girsanov transformation with kernel $S(t, T+\delta)$, we have

$$dW_t^{T+\delta} = dW_t - S(t, T+\delta)\ dt$$

for a $\mathbb{Q}$–Brownian motion W_t. Thus the riskneutral drift is directly determined by the volatility structure (as it is in the HJM model), giving riskneutral forward LIBOR rate dynamics

$$\frac{dL(t,T)}{L(t,T)} = -\lambda(t,T)\cdot S(t,T+\delta)\ dt + \lambda(t,T)\ dW_t$$

Now suppose that we want to create an HJM model in which forward LIBOR rates $L(t,T)$ have a deterministic volatility structure $\lambda(t,T)$. Above, we found that

$$S(t,T) - S(t,T+\delta) = \int_T^{T+\delta} \sigma(t,u)\ du = \frac{\delta L(t,T)}{1+\delta L(t,T)}\lambda(t,T)$$

(where S and σ are the bond and instantaneous forward rate volatilities respectively). In order to find the bond volatilities, it is necessary to impose some additional conditions. Set

$$\sigma(t,u) = 0 \quad \text{when } 0 \le u - t \le \delta$$

(This is the fundamental assumption made in BGM(1997)).
Now find the bond volatilities by a recursive procedure:

- Choose n such that $n\delta \le T - t < (n+1)\delta$. Equivalently $n = \sup\{k \in \mathbb{N} : k\delta \le T - t\} = [\delta^{-1}(T-t)]$ (where $[x]$ is the integer part of x).
- Then $S(t, T-n\delta) = -\int_t^{T-n\delta} \sigma(t,u)\ du = 0$, because $0 \le u - t \le \delta$ when $t \le u \le T - n\delta$.
- Thus

  $$\begin{aligned} S(t,T) = [S(t,T) - S(t,T-\delta)] + [S(t,T-\delta) - S(t,T-2\delta)] + \dots \\ \dots + [S(t,T-(n-1)\delta) - S(t,T-n\delta)] \end{aligned}$$

  implies

  $$\begin{aligned} S(t,T) = -\frac{\delta L(t,T-\delta)}{1+\delta L(t,T-\delta)}\lambda(t,T-\delta) - \frac{\delta L(t,T-2\delta)}{1+\delta L(t,T-2\delta)}\lambda(t,T-2\delta) - \dots \\ \dots - \frac{\delta L(t,T-n\delta)}{1+\delta L(t,T-n\delta)}\lambda(t,T-n\delta) \end{aligned}$$

- i.e.

  $$S(t,T) = -\sum_{k=1}^{[\delta^{-1}(T-t)]} \frac{\delta L(t,T-k\delta)}{1+\delta L(t,T-k\delta)}\lambda(t,T-k\delta)$$

Equivalently,

(i) Define $S(t,T) = 0$ for $0 \le T - t < \delta$.

(ii) Then define $S(t,T) = S(t,T-\delta) - \frac{\delta L(t,T-\delta)}{1+\delta L(t,T-\delta)}\lambda(t,T-\delta)$ for $\delta \le T-t < 2\delta$.
(Note that if $\delta \le T-t < 2\delta$, then $0 \le (T-\delta)-t < \delta$, so $S(t,T-\delta)$ has already been defined.)

(iii) Then define $S(t,T) = S(t,T-\delta) - \frac{\delta L(t,T-\delta)}{1+\delta L(t,T-\delta)}\lambda(t,T-\delta)$ for $2\delta \le T-t < 3\delta$.
(Note that if $2\delta \le T-t < 3\delta$, then $\delta \le (T-\delta)-t < 2\delta$, so $S(t,T-\delta)$ has already been defined.)

(iv) ...etc.

In this way, if we specify bond volatilities by this *forward induction*, then we will have an HJM model in which the forward LIBOR rates $L(t,T)$ have the required deterministic volatilities $\lambda(t,T)$. Since each $L(t,T)$ is a strictly postive $\mathbb{Q}^{T+\delta}$–martingale, it follows that each $L(t,T)$ is lognormal under $\mathbb{Q}^{T+\delta}$, and thus that the Black formula for caps is valid in this model.

8.6.2 The Musiela–Rutkowski Approach to Forward LIBOR

Unlike the BGM–approach, which lies within the HJM framework and specifies a model of forward LIBOR rates $L(t,T)$ for all maturities T (below the horizon T^*), the Musiela Rutkowski (MR) approach only specifies LIBOR rates for a discrete set of maturities. We start with a discrete tenor structure

$$0 < T_0 < T_1 < \cdots < T_N = T^* \qquad \delta_n = T_n - T_{n-1}$$

and define $T_{-1} = 0$ (for ease of handling certain formulas). We further assume that we are *given*

- A family of bounded adapted processes $\lambda(t,T_n)$ for $n = 0,\ldots,N-1$ which represent the volatilities of the forward LIBOR rates $L(t,T_n)$.
- An initial term structure $P(0,T_n)$ of discount bond prices (used to specify the initial conditions of the SDE's which we will write down for the LIBOR rates). We further assume that $P(0,T_0) > P(0,T_1)\cdots > P(0,T_N)$.

In contrast to the BGM approach, we do not need a bond price dynamics at all, i.e. we will attempt to model LIBOR rates directly.

Before we construct the MR model of LIBOR rates, a lemma which will prove useful

Lemma 8.6.1 *If X, Y are adapted processes*

$$dX_t = \alpha_t\, dW_t \qquad dY_t = \beta_t\, dW_t$$

and if $Z_t = \frac{1}{1+Y_t}$, then

$$\begin{aligned} d(Z_tX_t) &= Z_t(\alpha_t - \beta_t Z_t X_t)\cdot(dW_t - \beta_t Z_t\, dt) \\ i.e.\quad d(Z_tX_t) &= \eta_t \cdot (dW_t - \beta_t Z_t\, dt) \end{aligned}$$

for some process η_t.

Proof: A straightforward application of Itô's formula.

□

Whereas the BGM approach shows how to define bond volatilities by *forward induction*, the MR approach directly constructs a set of measures under which forward LIBOR rates have the required volatility structure by *backward induction*. It is therefore convenient to introduce the following backward notation. Put

$$T_k^* = T_{N-k} \quad \text{so that} \quad T^* = T_0^* > T_1^* > \cdots > T_N^* = T_0$$

We start by working under a T_N–forward riskneutral measure $\mathbb{Q}^{T_N} = \mathbb{Q}^{T_0^*}$, together with a $\mathbb{Q}^{T_N}$–Brownian motion $W^{T_N} = W_t^{T_0^*}$. it is not necessary to construct this measure: we can assume that $\mathbb{Q}^{T_N}$ is the measure $\mathbb{P}$ which governs our model, and that W^{T_N} is the W_t which drives the economy. Ultimately, we will be able to specify all the dynamics under this measure, the *terminal measure*. Let $L(t, T_1^*) = L(t, T_{N-1})$ be a process which satisfies the SDE plus initial value

$$dL(t, T_1^*) = L(t, T_1^*)\lambda(t, T_1^*)\, dW_t^{T_N}$$
$$L(0, T_1^*) = \frac{P(0, T_1^*) - P(0, T_0^*)}{\delta_N P(0, T_0^*)}$$

This *defines* the forward LIBOR rate $L(t, T_1^*) = L(t, T_{N-1})$ in the MR model.
We now use this to define the forward LIBOR rate $L(t, T_2^*) = L(t, T_{N-2})$. To do so, we need to construct the forward riskneutral measure for maturity T_2^*. Under $\mathbb{Q}^{T_2^*}$, all the bond ratios $\frac{P(t,T_n^*)}{P(t,T_2^*)}$ are martingales. Now define the ratio

$$U_{N-n+1}(t, T_k) = \frac{P(t, T_k)}{P(t, T_n)} \quad \text{or, equivalently} \quad U_n(t, T_k^*) = \frac{P(t, T_k^*)}{P(t, T_{n-1}^*)}$$

and note that each $U_n(t, T_k^*)$ is required to be a martingale under the measure $\mathbb{Q}^{T_{n-1}^*}$ (which we must still construct). Further note that

$$U_2(t, T_k^*) = \frac{U_1(t, T_k^*)}{1 + \delta_N L(t, T_1^*)}$$

so that by the lemma,

$$dU_2(t, T_k^*) = \eta_{k,t} \cdot \left(dW^{T_N} - \frac{\delta_N L(t, T_1^*)}{1 + \delta_N L(t, T_1^*)}\lambda(t, T_1^*)\, dt \right)$$

for some process $\eta_{k,t}$ (whose exact nature is not important right now). In order for each $U_2(t, T_k^*)$ to be a martingale, it suffices to find a measure under which

$$W_t^{T_1^*} = W_t^{T_{N-1}} = W_t^{T_N} - \int_0^t \frac{\delta_N L(s, T_1^*)}{1 + \delta_N L(s, T_1^*)}\lambda(s, T_1^*)\, ds$$

is a Brownian motion. This is possible if we perform a Girsanov transformation from $\mathbb{Q}^{T_N} = \mathbb{Q}^{T_0^*}$ with kernel $\gamma(s, T_1^*) = \frac{\delta_N L(s,T_1^*)}{1+\delta_N L(s,T_1^*)}\lambda(s, T_1^*)$, i.e. if we define

$$\frac{d\mathbb{Q}^{T_1^*}}{d\mathbb{Q}^{T_0^*}} = \mathcal{E}_{T_1^*}\left(\int_0^\cdot \gamma(s, T_1^*)\, dW^{T_0^*} \right)$$

We now let $L(t, T_2^*)$ be a process which solves the SDE and initial condition

$$dL(t, T_2^*) = L(t, T_2^*)\lambda(t, T_2^*)\, dW_t^{T_1^*}$$
$$L(0, T_2^*) = \frac{P(0, T_2^*) - P(0, T_1^*)}{\delta_{N-1} P(0, T_1^*)}$$

We continue in this way: Suppose that we have already constructed the LIBOR rate processes $L(t, T_1^*), \dots, L(t, T_n^*)$, for $n < N - 1$. Suppose further that this has been done so that each forward measure and Brownian motion has been specified, in particular that we have already constructed $\mathbb{Q}^{T_{n-1}^*}$ and $W_t^{T_{n-1}^*}$, and that $dL(t, T_n^*) = L(t, T_n^*)\lambda(t, T_n^*)\, dW_t^{T_{n-1}^*}$ under $\mathbb{Q}^{T_{n-1}^*}$. We must now construct a measure $\mathbb{Q}^{T_n^*}$ and an associated Brownian motion $W_t^{T_n^*}$. We require that each $U_{n+1}(t, T_k^*)$ is a $\mathbb{Q}^{T_n^*}$–martingale. Now

$$U_{n+1}(t, T_k^*) = \frac{U_n(t, T_k^*)}{1 + \delta_{N-n+1} L(t, T_n^*)}$$

Using the lemma, we see that

$$dU_{n+1}(t, T_k^*) = \eta_{k,t} \cdot \left(dW^{T_{n-1}^*} - \frac{\delta_{N-n+1} L(t, T_n^*)}{1 + \delta_{N-n+1} L(t, T_n^*)} \lambda(t, T_n^*)\, dt \right)$$

for some process $\eta_{k,t}$ (whose exact nature is not important right now). In order for each $U_{n+1}(t, T_k^*)$ to be a martingale, it suffices to find a measure under which

$$W_t^{T_n^*} = W_t^{T_{n-1}^*} - \int_0^t \frac{\delta_{N-n+1} L(s, T_n^*)}{1 + \delta_{N-n+1} L(s, T_n^*)} \lambda(s, T_n^*)\, ds$$

is a Brownian motion. This is possible if we perform a Girsanov transformation from $\mathbb{Q}^{T_{n-1}^*}$ with kernel $\gamma(s, T_n^*) = \frac{\delta_{N-n+1} L(s, T_n^*)}{1+\delta_{N-n+1} L(s, T_n^*)} \lambda(s, T_n^*)$, i.e. if we define

$$\frac{d\mathbb{Q}^{T_n^*}}{d\mathbb{Q}^{T_{n-1}^*}} = \mathcal{E}_{T_n^*} \left(\int_0^\cdot \gamma(s, T_n^*)\, dW_t^{T_{n-1}^*} \right)$$

We now let $L(t, T_{n+1}^*)$ be a process which solves the SDE and initial condition

$$dL(t, T_{n+1}^*) = L(t, T_{n+1}^*)\lambda(t, T_{n+1}^*)\, dW_t^{T_n^*}$$
$$L(0, T_{n+1}^*) = \frac{P(0, T_{n+1}^*) - P(0, T_n^*)}{\delta_{N-n} P(0, T_n^*)}$$

We have now constructed a sequence of processes $L(t, T_n)$ which are models of the forward LIBOR rates, with the desired volatilities. Since we also know the Girsanov kernels of each transformation, we can specify all LIBOR rate dynamics under the terminal measure. Inductively,

$$\begin{aligned}
dL(t, T_n^*) &= L(t, T_n^*)\lambda(t, T_n^*)\, dW_t^{T_{n-1}^*} \\
&= -L(t, T_n^*)\lambda(t, T_n^*)\gamma(t, T_{n-1}^*)\, dt + L(t, T_n^*)\lambda(t, T_n^*)\, dW_t^{T_{n-2}^*} \\
&= -L(t, T_n^*)\lambda(t, T_n^*)[\gamma(t, T_{n-1}^*) + \gamma(t, T_{n-2}^*)]\, dt + L(t, T_n^*)\lambda(t, T_n^*)\, dW_t^{T_{n-3}^*} \\
&= \dots \\
&= -L(t, T_n^*)\lambda(t, T_n^*) \sum_{k=1}^{n-1} \gamma(t, T_{n-k}^*)\, dt + L(t, T_n^*)\lambda(t, T_n^*)\, dW_t^{T_0^*}
\end{aligned}$$

where

$$\gamma(t,T_k^*) = \frac{\delta_{N-k+1}L(t,T_k^*)}{1+\delta_{N-k+1}L(t,T_k^*)}\lambda(t,T_k^*)$$

and hence, when we translate from backwards time to ordinary time,

> The Musiela–Rutkowski forward LIBOR rate dynamics under the terminal measure $\mathbb{Q}^{T_N}$ are given by
>
> $$dL(t,T_n) = -L(t,T_n)\lambda(t,T_n)\sum_{k=n+1}^{N-1}\frac{\delta_{k+1}\lambda(t,T_k)L(t,T_k)}{1+\delta_{k+1}L(t,T_k)}\,dt + L(t,T_n)\lambda(t,T_n)\,dW_t^{T_N}$$
>
> This must be solved recursively: First find the solution for $L(t,T_{N-1})$. Once this has been found, find the solution for $L(t,T_{N-2})$. Note that the SDE for $L(t,T_{N-2})$ also contains $L(t,T_{N-1})$, but we've already found that. Then solve the SDE for $L(t,T_{N-3})$ (which contains $L(t,T_{n-1})$ and $L(t,T_{N-2})$; these have been determined). And so on...

It is therefore *possible* to find a model in which LIBOR rates have the required volatilities $\lambda(t,T_n)$. If these volatilities are deterministic, then each $L(t,T_n)$ will be lognormal under $\mathbb{Q}^{T_{n+1}}$. In that case, the Black formula for caps will be exact.

8.6.3 Jamshidian's Approach to Forward LIBOR

Like the Musiela–Rutkowski approach, Jamshidian(1997) does not require bond price dynamics, and models LIBOR rates for a discrete set of tenor dates $0 = T_{-1} < T_0 < T_1 < \cdots < T_N = T^*$ via a backward induction. But instead of working under the terminal measure, Jamshidian defines a *spot LIBOR measure*. This measure is obtained if we take as numéraire a certain portfolio of zero coupon bonds with unit initial value.

We begin by observing that the prices of discount bonds are not completely determined by the forward LIBOR rates. This is true at tenor dates, but if t lies between tenor dates, e.g. $T_n < t < T_{n+1}$, then $P(t,T_{n+k}) = P(t,T_{n+1})\cdot\frac{1}{1+\delta_{n+2}L(t,T_{n+1})}\cdot\cdots\cdot\frac{1}{1+\delta_{n+k}L(t,T_{n+k-1})}$. Thus knowledge of the LIBOR rates is not enough — we also have to know the discount factor to the next tenor date (i.e. $P(t,T_{n+1})$). By working under the spot LIBOR measure, this problem can be circumvented.

Consider the following portfolio of discount bonds X. Its initial value is \$1.00. At all subsequent times, all wealth is invested in the next–to–mature bond. Thus at $t=0$, \$1.00 is invested in $P(t,T_0)$. At T_0, the payoff of these bonds is reinvested in $P(t,T_1)$ and at T_1, the payoff is reinvested in $P(t,T_2)$, etc. Thus at time T_n, the value of the portfolio is

$$\begin{aligned} X_{T_n} &= \frac{P(T_n,T_{n+1})}{P(0,T_0)\cdot P(T_0,T_1)\cdot\cdots\cdot P(T_n,P(T_{n+1})} \\ &= \text{value of } T_{n+1}\text{–bonds} \times \text{ no. of } T_{n+1}\text{–bonds} \end{aligned}$$

An instant later, when $T_n \le t < T_{n+1}$, the value X_t of the portfolio is simply

$$X_t = \frac{P(t,T_{n+1})}{P(0,T_0)\cdot P(T_0,T_1)\cdot\cdots\cdot P(T_n,P(T_{n+1})}$$

because the value of the T_{n+1}–bond has changed, but the number of T_{n+1}–bonds in the portfolio has not. Hence

$$X_t = P(t, T_{n(t)}) \cdot \prod_{k=0}^{n(t)} P^{-1}(T_{k-1}, T_k) \qquad (*)$$
$$\text{where } n(t) = \inf\{n : T_n > t\}$$

A *spot LIBOR measure* $\mathbb{Q}_X$ is obtained by taking X_t as numéraire, so that each asset ratio process $\frac{P(t,T_n)}{X_t}$ is a $\mathbb{Q}_X$–martingale. The asset ratios can be written as

$$\begin{aligned}
\frac{P(t, T_{n+1})}{X_t} &= \frac{P(t, T_{n(t)}) \prod_{k=n(t)+1}^{n} (1 + \delta_k L(t, T_{k-1}))^{-1}}{P(t, T_{n(t)}) \prod_{k=0}^{n(t)} (1 + \delta_k L(T_{k-1}, T_{k-1})} \\
&= \prod_{k=0}^{n(t)} (1 + \delta_k L(T_{k-1}, T_{k-1}))^{-1} \prod_{k=n(t)+1}^{n} (1 + \delta_k L(t, T_{k-1}))^{-1} \\
&= \prod_{k=0}^{n} (1 + \delta_k L(t \wedge T_{k-1}, T_{k-1}))^{-1}
\end{aligned}$$

Hence the prices of the asset ratios are completely determined by the LIBOR processes. We now aim to describe the LIBOR rate dynamics under the spot LIBOR measure $\mathbb{Q}_X$, and that this requires knowledge only of the LIBOR rate volatilities (and not, say, bond or instantaneous forward rate volatilities as well). For the moment, assume that bond price dynamics are given by some Itô processes

$$\frac{dP(t, T_n)}{P(t, T_n)} = m(t, T_n)\, dt + S(t, T_n)\, dW_t$$

under the "real–world" probability measure $\mathbb{P}$. By definition of X_t (i.e. by $(*)$), we see that

$$\frac{dX_t}{X_t} = m(t, T_{n(t)})\, dt + S(t, T_{n(t)})\, dW_t$$

Moreover, if we apply Itô's formula to $1 + \delta_{n+1} L(t, T_n) = \frac{P(t,T_n)}{P(t,T_{n+1})}$, we see that

$$\begin{aligned}
dL(t, T_n) = \frac{P(t, T_n)}{\delta_{n+1} P(t, T_{n+1})} \Big[&\Big(m(t, T_n) - m(t, T_{n+1}) - (S(t, T_n) - S(t, T_{n+1})) S(t, T_{n+1}) \Big)\, dt \\
&- \Big(S(t, T_n) - S(t, T_{n+1}) \Big)\, dW_t \Big] \\
= \mu(t, T_n)\, dt + \zeta(t, T_n)\, dW_t &
\end{aligned}$$

where

$$\begin{aligned}
\mu(t, T_n) &= \frac{P(t, T_n)}{\delta_{n+1} P(t, T_{n+1})} \Big(m(t, T_n) - m(t, T_{n+1}) \Big) - \zeta(t, T_n) S(t, T_{n+1}) \\
\zeta(t, T_n) &= \frac{P(t, T_n)}{\delta_{n+1} P(t, T_{n+1})} \Big(S(t, T_n) - S(t, T_{n+1}) \Big)
\end{aligned}$$

It follows that

$$S(t, T_{n(t)}) - S(t, T_{j+1}) = \sum_{k=n(t)}^{j} \frac{\delta_{k+1} \zeta(t, T_k)}{1 + \delta_{k+1} L(t, T_k)} \qquad (**)$$

for $j \geq n(t)$.

If $\gamma(t)$ is the Girsanov kernel for transforming $\mathbb{P}$ to $\mathbb{Q}_X$, i.e. if $\frac{d\mathbb{Q}_X}{d\mathbb{P}} = \mathcal{E}_{T^*}\left(\int_0^\cdot \gamma_t \, dW_t\right)$, then $\frac{P(t,T_n)}{X_t}$ has zero drift under $\mathbb{Q}_X$. But

$$d\frac{P(t,T_n)}{X_t} = \frac{P(t,T_n)}{X_t}\bigg[\Big(m(t,T_n) - m(t,T_{n(t)}) - S(t,T_{n(t)}) \cdot (S(t,T_n) - S(t,T_{n(t)}))\Big)\, dt \\ + \Big(S(t,T_n) - S(t,T_{n(t)})\Big)\, dW_t\bigg]$$

Now in the Girsanov transformation, $(S(t,T_n) - S(t,T_{n(t)})) \cdot \gamma_t$ is added to the $\mathbb{P}$–drift to obtain the $\mathbb{Q}_X$–drift, which is zero, and so

$$m(t,T_n) - m(t,T_{n(t)}) - S(t,T_{n(t)}) \cdot (S(t,T_n) - S(t,T_{n(t)})) + (S(t,T_n) - S(t,T_{n(t)})) \cdot \gamma_t = 0$$

which yields

$$m(t,T_n) - m(t,T_{n+1}) = \Big(S(t,T_{n(t)}) - \gamma_t\Big) \cdot \Big(S(t,T_n) - S(t,T_{n(t)})\Big)$$

for $n = 0, \ldots, N$. It follows that

$$\begin{aligned} m(t,T_n) - m(t,T_{n+1}) &= \Big(m(t,T_n) - m(t,T_{n(t)}\Big) - \Big(m(t,T_{n+1}) - m(t,T_{n(t)})\Big) \\ &= \Big(S(t,T_{n(t)}) - \gamma_t\Big) \cdot \Big(S(t,T_n) - S(t,T_{n+1})\Big) \end{aligned}$$

Now multiply both sides of this equation by $\frac{P(t,T_n)}{\delta_{n+1}P(t,T_{n+1})}$ to obtain

$$\frac{P(t,T_n)}{\delta_{n+1}P(t,T_{n+1})}\Big(m(t,T_n) - m(t,T_{n+1})\Big) = \zeta(t,T_n)\Big(S(t,T_{n(t)}) - \gamma_t\Big)$$

Looking back to the definitions of μ and ζ in the dynamics of $L(t,T_n)$, we see that

$$\mu(t,T_n) = \zeta(t,T_n)\Big(S(t,T_{n(t)}) - \gamma_t - S(t,T_{n+1})\Big)$$

and hence

$$dL(t,T_n) = \zeta(t,T_n) \cdot \bigg[\Big(S(t,T_{n(t)}) - S(t,T_{n+1}) - \gamma_t\Big)\, dt + dW_t\bigg]$$

These are, of course, the $\mathbb{P}$–dynamics. To get the $\mathbb{Q}_X$–dynamics, we must add volatility $\times$ kernel $= \zeta_t \cdot \gamma_t$ to the drift to obtain

$$dL(t,T_n) = \zeta(t,T_n) \cdot \bigg[\Big(S(t,T_{n(t)}) - S(t,T_{n+1})\Big)\, dt + dW_t^X\bigg]$$

where $W_t^X = W_t - \int_0^t \gamma_u \, du$ is a $\mathbb{Q}_X$–Brownian motion. Finally, using $(**)$, we obtain

$$dL(t,T_n) = \sum_{k=n(t)}^{n} \frac{\delta_{k+1}\zeta(t,T_k) \cdot \zeta(t,T_n)}{1 + \delta_{k+1}L(t,T_k)}\, dt + \zeta(t,T_n) dW_t^X$$

These are the forward LIBOR rate dynamics under the spot LIBOR measure.

8.7 Exercises

1. An endowment option X is a very long term European call option. Typically,

   - At issue, the initial strike K_0 is set to approximately 50% of the current stock price.
   - The options are inflation and dividend protected:
     - The strike price increases at the short term riskless rate.
     - The strike price is decreased by the size of the dividend each time a dividend is paid.
   - The payoff at expiry T is $X_T = (S_T - K_T)^+$.

   We will make the simplifying assumption that the stock pays no dividends. This can be accomplished by regarding the stock price as the theoretical price of a mutual fund which starts off at one share, and reinvests all dividends in that share. We have, in the risk–neutral world,

   $$dS_t = r_t S_t\, dt + \sigma_t S_t\, dW_t \quad dA_t = r_t A_t\, dt$$

   where S is the share and A the money market account (with $A_0 = 1$). Clearly $K_t = K_0 A_t$. By changing the numéraire to A_t, show that, when the volatility σ_t is deterministic,

   $$X_0 = S_0 N(d_+) - K_0 N(d_-)$$

   where

   $$d_\pm = \frac{\ln \frac{S_0}{K_0} \pm \frac{1}{2}\sigma_{av}^2 T}{\sigma_{av}\sqrt{T}} \quad \text{and} \quad \sigma_{av}^2 = \frac{1}{T}\int_0^T \sigma_t^2\, dt$$

2. Use the change–of–numéraire technique to show how to calculate the value of an option which pays the minimum of two assets S^1, S^2. Assume that the "real world" dynamics of the assets are Itô diffusions of the form

   $$dS_t^i = S_t^i[\mu_i\, dt + \hat{\sigma}_i\, d\hat{W}_t^i]$$

   where μ_i, σ_i are constants, and that the correlation of returns is a constant ρ. Further assume that S_t^i has a continuously paid dividend with constant dividend yield q^i.

3. Consider a European call C on share S traded on FTSE. S_t and C are priced in pounds, but the strike of the call is in dollars. Initially, the option is at–the–money. The dollar strike does not change, but because exchange rates are not fixed, the pound strike does. Let X_t be the $\frac{\text{dollar}}{\text{pound}}$-rate, Y_t the $\frac{\text{pound}}{\text{dollar}}$–rate. Assume dynamics

   $$\begin{aligned} dS_t &= \alpha_S S_t\, dt + \delta_S S_t\, dW_t^S \\ dX_t &= \alpha_X X_t\, dt + \delta_X X_t\, dW_t^X \\ dY_t &= \alpha_Y Y_t\, dt + \delta_Y Y_t\, dW_t^Y \end{aligned}$$

   where W^S, W^X, W^Y are correlated Brownian motions.

   3.1 Apply Itô's formula to show

   $$dY_t = \alpha_Y Y_t\, dt + \delta_Y Y_t(-dW_t^X) \qquad \alpha_Y = -\alpha_X + \delta_X^2, \quad \delta_Y = \delta_X$$

3.2 Let ρ be the correlation between W^S and W^X. Let $W_t = (W_t^1, W_t^2)$ be a two–dimensional standard Brownian motion, and rewrite the above dynamics

$$\begin{aligned} dS_t &= \alpha_S S_t \, dt + S_t \sigma_S \, dW_t \\ dX_t &= \alpha_X X_t \, dt + X_t \sigma_X \, dW_t \\ dY_t &= \alpha_Y Y_t \, dt + Y_t \sigma_Y \, dW_t \end{aligned}$$

Show that we must have

$$\begin{aligned} ||\sigma_X||^2 = \delta_X^2 \quad ||\sigma_Y||^2 = \delta_Y^2 \quad ||\sigma_S||^2 = \delta_S^2 \\ \sigma_X \cdot \sigma_S = \rho\delta_X\delta_S \quad \sigma_Y \cdot \sigma_S = -\rho\delta_X\delta_S \end{aligned}$$

3.3 The initial pound strike is $K_0^p = S_0$, and the initial dollar strike is $K^d = S_0 X_0$ (at–the–money), which remains fixed. At maturity, the pound strike is $K_T = K^d Y_T$. Define $S_t^d = S_t X_t$ to be the dollar price of S at time t. Show that

$$dS_t^d = S_t^d[\alpha_S + \alpha_X + \sigma_S \cdot \sigma_X] \, dt + S_t^d(\sigma_S + \sigma_X) \, dW_t$$

3.4 Now convert this to a system with a one–dimensional Brownian motion V_t:

$$dS_t^d = S_t^d[\alpha_S + \alpha_X + \sigma_S \cdot \sigma_X] \, dt + S_t^d \delta_{S^d} \, dV_t$$

where

$$\delta_{S^d}^2 = ||\sigma_X + \sigma_S||^2 = (\delta_X^2 + \delta_S^2 + 2\rho\delta_X\delta_S)$$

3.5 Now we have a plain vanilla call on an asset S^d with (fixed) strike K^d. Find the dollar price C_t^d of this option:

$$C_t^d = S^d N(d_+) - e^{-r_d(T-t)} K^d N(d_-)$$

where where r_d is the riskless *dollar* rate, and

$$d_\pm = \frac{\ln \frac{S_t^d}{S_0^d} + (r_d \pm \frac{1}{2}\delta_{S^d}^2)(T-t)}{\delta_{S^d}\sqrt{T-t}}$$

3.6 Conclude that the pound price of the option is

$$C_t = S_t N(d_+) - e^{-r_d(T-t)} \tfrac{S_0 X_0}{X_t} N(d_-) \qquad d_\pm = \frac{\ln \frac{S_t X_t}{S_0 X_0} + (r_d \pm \frac{1}{2}(\delta_X^2 + \delta_S^2 + 2\rho\delta_X\delta_S))(T-t)}{\sqrt{(\delta_X^2 + \delta_S^2 + 2\rho\delta_X\delta_S)(T-t)}}$$

3.7 If we had tried to price the option directly in pounds, we would have had (explain this)

$$C_T = (S_T - S_0(Y_T/Y_0))^+$$

Very naturally, we would have considered the numeraire Y_t. This would have been a mistake, for although Y_t is a traded asset (namely the pound price of a dollar note), this is not a non–dividend paying asset: Y_t has a continuous dividend yield equal to the riskless dollar rate r_d. Thus discounted Y_t is not a $\mathbb{Q}$–martingale. Instead, therefore,

consider the process $\hat{Y}_t = Y_t e^{r_d t}$. (i.e. all dividends = interest reinvested in the dollar money market account). Show that

$$C_t = \hat{Y}_t \mathbb{E}_{\hat{\mathbb{Q}}}[(\hat{S}_t - K')^+ | \mathcal{F}_t]$$

where $K' = e^{-r_d T} S_0 / Y_0$ and $\mathbb{Q}$ is the equivalent martingale measure associated with $\hat{Y}$.

3.8 Find the $\hat{\mathbb{Q}}$–dynamics of $\hat{S}_t$ (with a two–dimensional standard Brownian motion).

3.9 Convert this to S_t–dynamics with a one–dimensional Brownian motion.

3.10 Hence show that

$$C_t \hat{Y}_t [\hat{S}_t N(d_+) - K' N(d_-)]$$

where

$$d_\pm = \frac{\ln \frac{\hat{S}_t}{K'} \pm \frac{1}{2}\hat{\delta}^2 (T-t)}{\hat{\delta}\sqrt{T-t}}$$

and $\hat{\delta} = ||\sigma_S - \sigma_Y|| = \sqrt{\delta_S^2 + \delta_Y^2 + 2\rho\delta_S\delta_Y}$.

3.11 Finally show that this coincides with the formula obtained earlier.
In this case, you see that it is slightly easier to value the option in dollars than it is in pounds.

4. Suppose the bond price dynamics are given by

$$dp(t,T) = p(t,T)M(t,T)\,dt + p(t,T)v(t,T)\,dW_t$$

Show that in that case the forward rate dynamics are given by

$$df(t,T) = \alpha(t,T)\,dt + \sigma(t,T)\,dW_t$$

where

$$\alpha(t,T) = v_T(t,T)v(t,T) - m_T(t,T) \qquad \sigma(t,T) = -v_T(t,T)$$

[Hint: Apply Itô's formula to $\ln p(t,T)$, write this in integrated form, and differentiate with respect to T.]

5. Let $\{y(0,T) : T \geq 0\}$ denote the zero–coupon yield curve at $t = 0$. Assume that, apart from the zero coupon bonds, we also have exactly one fixed coupon bond for every maturity T. enote the yield–to–maturity of the fixed coupon bond by $y_M(0,T)$. We now have 3 curves to consider, the forward rate curve $f(0,T)$, the zero yield curve $y(0,T)$ and the coupon yield curve $y_M(0,T)$.

5.1 Show that $f(0,T) = y(0,T) + T\dfrac{\partial y(0,T)}{\partial T}$

5.2 Assume that the zero yield curve is an increasing function of T. Show that in that case

$$y_M(0,T) \leq y(0,T) \leq f(0,T)$$

for all T. Show that the inequalities are reversed if the zero yield curve is decreasing. Explain this phenomenon in terms of simple economics.

5.3 Yield curves can be both upward and downward sloping. Can this be true for bond price curves $p(0,T)$?

6. In the Cox–Ingersoll–Ross model, the risk–neutral short rate dynamics assumed are

$$dr_t = (b - ar_t)\,dt + \sigma\sqrt{r_t}\,dW_t, \qquad a, b, \sigma, r_0 > 0$$

6.1 Explain (heuristically) why this process is mean–reverting and non–negative.

6.2 This is an affine short rate model. By plugging $p(t,T) = e^{A(t,T) - B(t,T)r_t}$ into the term structure PDE, show that we obtain two coupled ODE's

$$\begin{aligned} B_t &= aB + \frac{1}{2}\sigma^2 B^2 - 1 \qquad && B(T,T) = 0 \\ A_t &= bB && A(T,T) = 0 \end{aligned}$$

6.3 To solve the Riccati equation for B, try a solution of the form

$$B(t,T) = \frac{X(t)}{cX(t) + d}$$

Choose c to ensure that $a + \frac{1}{2}\sigma^2 - c^2 = 0$. Show that we then obtain a order linear differential equation

$$X_t + \kappa X = -d \qquad \text{where } \kappa = -a + 2c = \sqrt{a^2 + 2\sigma^2}$$

6.4 Solve the ODE to obtain

$$X(t) = \frac{d}{\kappa}[e^{\kappa(T-t)} - 1]$$

6.5 Hence show that

$$B(t,T) = \frac{2(e^{\kappa(T-t)} - 1)}{2\kappa + (a+\kappa)(e^{\kappa(T-t)} - 1)} \qquad \text{where } \kappa = \sqrt{a^2 + 2\sigma^2}$$

6.6 Verify by differentiation that

$$A(t,T) = \frac{2b}{\sigma^2} \ln\left[\frac{2\kappa e^{\frac{1}{2}(a+\kappa)(T-t)}}{2\kappa + (a+\kappa)(e^{\kappa(T-t)} - 1)}\right]$$

7. 7.1 Show that the Hull–White model $dr = (\theta(t) - ar)\,dt + \sigma\,dW_t$ is obtained if one starts with a HJM model given by

$$df(t,T) = \alpha(t,T)\,dt + \sigma e^{-a(T-t)}\,dW_t$$

Hence compute the function $\theta(t)$ which will make the short rate model fit the initial term structure:

$$\theta(t) = f_T^*(0,t) + af^*(0,t) + \frac{\sigma^2}{a}[1 - e^{-2at}]$$

where $\{f^*(0,T) : T \geq 0\}$ is the observed term structure of forward rates. It follows that the Hull–White model can also be fitted to any initial term structure. What is the distribution of the forward rate $f(t,T)$?

7.2 Show that bond prices in the Hull–White model, fitted to the initial term structure, are given by

$$p(t,T) = \frac{p(0,T)}{p(0,t)} \exp\left(f(0,t)B(t,T) - \frac{\sigma^2}{4a}B^2(t,T)(1-e^{-2at}) - B(t,T)r_t \right)$$

where $B(t,T) = \frac{1}{a}[1-e^{-a(T-t)}]$.
[Hint: The Hull–White model is an affine term structure model, i.e. $p(t,T) = e^{A(t,T)-B(t,T)r_t}$. $B(t,T)$ is readily calculated. We can now find

$$A(t,T) = \int_t^T -\theta(u)B(u,T)\, du + \frac{1}{2}\sigma^2 B^2(0,t)$$

where $\theta(t)$ is as in (a), i.e. $b(u) = e^{-au}\frac{d}{du}x(u)e^{au}$, where $x(t) = e^{-at}r_0 + \int_0^t e^{-a(t-u)}\, du$. Integrating by parts leads to

$$\begin{aligned} A(t,T) &= f(0,t)B(t,T) + \ln\frac{p(0,T)}{p(0,t)} \\ &+ \frac{\sigma^2}{2}\left[\int_t^T b^2(u,T) - B^2(0,u)\, du + B^2(0,t)B(t,T)\right] \end{aligned}$$

which simplifies to give the required result.]

7.3 Show that the Hull–White price of a call C with strike K and maturity T on a bond $p(0,S)$ (where $S > T$) is given by

$$p(0,S)N(d_+) - Kp(0,T)N(d_-)$$

where

$$d_\pm = \frac{\ln\frac{p(0,S)}{Kp(0,T)} \pm \frac{1}{2}\sigma_{av}^2 T}{\sigma_{av}\sqrt{T}}$$

and where

$$\sigma_{av}^2 T = \int_0^T \frac{\sigma^2}{a^2}(e^{-aS} - e^{-aT})^2 e^{2at}\, dt = \frac{\sigma^2}{2a^3}(1-e^{-a(S-T)})^2(1-e^{-2aT})$$

8. Consider the domestic and the foreign bond market, with bond prices denoted by $p_d(t,T)$ and $p_f(t,T)$ respectively. Take as given a standard HJM model for the domestic forward rates $f_d(t,T)$

$$df_d(t,T) = \alpha_d(t,T)\, dt + \sigma_d(t,T)\, dW_t$$

where W_t is a multidimensional Brownian motion under the *domestic* martingale measure $\mathbb{Q}$. The foreign forward rates are denoted by $f_f(t,T)$, and their $\mathbb{Q}$–dynamics are given by

$$df_f(t,T) = \alpha_f(t,T)\, dt + \sigma_f(t,T)\, dW_t$$

Note that the same Brownian motion drives both bond markets. The exchange rate X (in units of domestic currency per unit of foreign currency) has $\mathbb{Q}$–dynamics

$$dX_t = \mu(t)X(t)\, dt + \sigma_X(t)X(t)\, dW_t$$

Show that under the domestic martingale measure the foreign forward rates satisfy the modified HJM drift condition

$$\alpha_f(t,T) = \sigma_f(t,T) \left[\int_t^T \sigma_f^{\mathrm{tr}}(t,s)\, ds - \sigma_X^{\mathrm{tr}}(t) \right]$$

9. A common implementation of the HJM framework uses the following forward rate dynamics:

$$df(t,T) = \alpha(t,T)\, dt + (\sigma_1, \sigma_2 e^{-\frac{\lambda}{2}(T-t)}) \cdot (dW_1(t), dW_2(t))$$

where $\sigma_1, \sigma_2, \lambda$ are non–negative constants, W_1, W_2 are independent $\mathbb{Q}$–Brownian motions, and $\mathbb{Q}$ is the equivalent risk–neutral measure.

This is a two–factor model. The first factor $W_1(t)$ can be interpreted as a source of noise that lasts a long time, affecting all maturities equally. The second factor $W_2(t)$ affects short maturity forward rates more than the long term rates (why?), and thus adds some extra volatility to the short term rates.

9.1 Show that the HJM drift conditions imply that

$$\alpha(t,T) = \sigma_1^2(T-t) - \frac{2\sigma_2^2}{\lambda} e^{-\frac{\lambda}{2}(T-t)}(e^{-\frac{\lambda}{2}(T-t)} - 1)$$

9.2 Hence show that

$$\begin{aligned} f(t,T) =& f(0,T) + \sigma_1^2 t(T - t/2) - 2(\sigma_2/\lambda)^2 [e^{-\lambda T}(e^{\lambda t} - 1) - 2e^{-(\lambda/2)T}(e^{(\lambda/2)t} - 1)] \\ &+ \sigma_1 W_1(t) + \sigma_2 \int_0^t e^{-(\lambda/2)(T-u)}\, dW_2(u) \end{aligned}$$

9.3 Show that the spot rate follows the process

$$\begin{aligned} r(t) &= f(0,t) + \frac{1}{2}\sigma_1^2 t^2 - 2(\sigma_2/\lambda)^2 [1 - e^{-(\lambda/2)t}]^2 \\ &+ \sigma_1 W_1(t) + \sigma_2 e^{-(\lambda/2)t} \int_0^t e^{(\lambda/2)u}\, dW_2(u) \end{aligned}$$

9.4 Is the short rate a Markov process, a Gaussian process, a stationary process? Explain.

9.5 Calculate the price $C(t)$ of a call option on the zero coupon bond $p(t,T)$. Assume that the option has strike K and expiry τ, where $t \leq \tau \leq T$.
[Hint: Let $p(t,\tau)$ be the numeraire. You know the HJM dynamics of zero coupon bonds under $\mathbb{Q}$, so the dynamics of $p(t,T)/p(t,\tau)$ under the EMM for $p(t,\tau)$ should be easy to find. Of course, something is going to be lognormal. Now use the general option pricing formula.]

9.6 As a check, assume that $\sigma_1 = 0.2, \sigma_2 = 0.3$ and $\lambda = 2$. Calculate the value of a call option on a two–year zero coupon bond with strike 0.9 and expiry 1 year. Today's prices are $P(0,1) = 0.9, P(0,2) = 0.81$. I get 0.076 (but I could be wrong, of course).

10. Consider a convertible bond X which, at T_0, allows the owner to convert the bond to c shares S of common stock. The bond is a zero coupon bond with face value 1.00 and maturity $T_1 > T_0$. The aim of this problem is to find the price the convertible bond at

some future date $t \leq T_0$. We will model the short rate using Ho–Lee dynamics. Initially, the (instantaneous) forward rate curve is flat with $f(0,T) = r_0$ for all maturities T.

We work under a risk-neutral measure $\mathbb{Q}$ where the share has dynamics

$$dS_t = r(t)S_t\, dt + \sigma_S S_t\, dW_t$$

and the short rate has dynamics

$$dr(t) = \theta(t)\, dt + \sigma_r\, dW_t$$

Here W_t is a two–dimensional $\mathbb{Q}$–Brownian motion, and σ_S, σ_r are constant vectors.

10.1 Let $p(t,T)$ be a non–convertible zero coupon bond with face value 1.00 and maturity T years. Calculate the observed term structure of bond prices $\{p^*(0,T) : T \geq 0\}$.

10.2 Let $\hat{\mathbb{Q}}$ be the forward risk–neutral measure for maturity T_1 years (i.e. the EMM for numéraire $p(t,T_1)$). By decomposing the convertible bond into its option and bond parts, show that

$$X_0 = cp(0,T_1)\mathbb{E}_{\hat{\mathbb{Q}}}\left[(\hat{S}_{T_0} - \tfrac{1}{c})^+\right] + p(0,T_1)$$

where $\hat{S}_t = \frac{S_t}{p(t,T_1)}$.

10.3 The Ho–Lee model is an affine term structure model, i.e. bond prices are of the form

$$p(t,T) = e^{A(t,T) - B(t,T)r(t)}$$

By substituting this expression into the term structure PDE, show that

$$B(t,T) = T - t \qquad A(t,T) = \int_t^T \theta(u)(u-T)\, du + \frac{1}{6}\sigma_r^2(T-t)^3$$

10.4 Show that

$$\frac{dp(t,T)}{p(t,T)} = r(t)\, dt - \sigma_r(T-t)\, dW_t$$

10.5 Fit the Ho–Lee model to the initial term structure of forward rates: Show that

$$\theta(t) = \sigma_r^2 t$$

10.6 Hence show that

$$r(t) = r_0 + \frac{1}{2}\sigma_r^2 t^2 + \sigma_r W_t$$

10.7 Hence, using the known initial bond prices and short rate dynamics, show that future bond prices are given by

$$p(t,T) = e^{-\frac{1}{2}\sigma_r^2 t(T-t)^2 - r(t)(T-t)}$$

What is the distribution of $p(t,T)$ under $\mathbb{Q}$?

10.8 Show that $\hat{S}_t$ has dynamics

$$\frac{d\hat{S}_t}{\hat{S}_t} = (\sigma_S + \sigma_r(T_1 - t))\, d\hat{W}_t$$

under the measure $\hat{\mathbb{Q}}$, where $\hat{W}_t$ is a $\hat{\mathbb{Q}}$–Brownian motion.
What is the distribution of $\hat{S}_t$ under $\hat{\mathbb{Q}}$?

10.9 Deduce that

$$X_0 = cS_0 N(d_+) - p(0,T_1)N(d_-) + p(0,T_1)$$

where

$$d_\pm = \frac{\ln \frac{cS_0}{p(0,T_1)} \pm \frac{1}{2}\Sigma^2 T_0}{\sqrt{\Sigma^2 T_0}} \qquad \Sigma^2 = \frac{1}{T_0}\int_0^{T_1} ||\sigma_S + \sigma_r(T_1 - s)||^2 \, ds$$

11. The aim of this problem is to calculate the price of an in–arrears caplet in the Ho–Lee model, where the short rate has riskneutral dynamics

$$dr_t = \theta(t)\, dt + \sigma\, dW_t$$

Here, σ is a constant, and W_t is a 1–dimensional Brownian motion under the risk neutral measure $\mathbb{Q}$. The caplet has payoff

$$0.5 \max\{L - R_c, 0\}$$

at expiry = 1 year, where L is the 6–month spot LIBOR rate in 6 months' time, and R_c is the cap rate. Use the following data:

$P(0,T)$	$e^{-r_0 T}$
r_0	10%
R_c	12%
σ	10%

Here $P(0,T)$ is the default–free zero coupon bond with face value 1 and maturity T.

We proceed as follows: We first show that the caplet is equivalent to a portfolio of put options on zero coupon bonds. Then we recast the Ho–Lee model within the HJM framework in order to fit it to the observed (flat) term structure, and calculate the prices of zero coupon bonds. Finally, we calculate the prices of vanilla options on zero coupon bonds.

11.1 First show that a caplet can be regarded as a portfolio of 6–month put options on the 1-year zero:

$$\begin{aligned} \text{Caplet } =& (1 + R_c \Delta T) \text{ put options on } P(t, T_2) \text{ with strike } \frac{1}{1 + R_c \Delta T} \\ & \text{and maturity } T_1 \end{aligned}$$

where $T_1 = 0.5$, $\Delta T = 0.5$, and $T_2 = T_1 + \Delta T = 1$.

11.2 The Ho–Lee model is an *affine* short rate model, with bond prices of the form $P(t,T) = e^{A(t,T) - B(t,T) r_t}$. By substituting this form of $P(t,T)$ into the term structure PDE, show that

$$\begin{aligned} B(t,T) &= T - t \\ A(t,T) &= -\int_t^T \theta(u)(T-u)\, du + \frac{1}{6}\sigma^2 (T-t)^3 \end{aligned}$$

11.3 In order to fit the short rate model to the observed term structure, we recast it in the HJM framework. You may use the facts about the HJM model which are stated on the formula sheet.
Using the relation between forward rates and zero coupon bond prices and the value of $B(t, T)$, show that the instantaneous forward rate $f(t, T)$ has a constant "volatility" σ, i.e. that the forward rate dynamics are

$$df(t, T) = \alpha(t, T)\ dt + \sigma\ dW_t$$

for some function $\alpha(t, T)$.

11.4 Use the HJM drift conditions to show that $\alpha(t, T) = \sigma^2(T - t)$

11.5 Hence show that

$$r_t = r_0 + \tfrac{1}{2}\sigma^2 t^2 + \sigma W_t$$

and conclude that

$$dr_t = \sigma^2 t\ dt + \sigma\ dW_t$$

11.6 Next, show that

$$A(t, T) = -\frac{1}{2}\sigma^2 t(T - t)^2$$

and thus that zero coupon bond prices are given by

$$P(t, T) = e^{-\frac{1}{2}\sigma^2 t(T-t)^2 - (T-t)r_t}$$

11.7 Now that bond prices have been found, we will price bond options. Recall the general option formula stated in the Formula Sheet. Change the numeraire to the 6-month zero coupon bond $P(0, T_1)$. Let $\hat{P}_t = \frac{P(t,T_2)}{P(t,T_1)}$. Write down the dynamics of $\hat{P}_t$ under T_1–forward neutral measure $\mathbb{Q}_1$.

11.8 Show that $\hat{P}_{T_1}$ is lognormally distributed under $\mathbb{Q}_1$, and find its distribution.

11.9 Similarly, find the dynamics of $\check{P}_t = \frac{1}{\hat{P}_t}$ under the T_2–forward measure $\mathbb{Q}_2$. Show that $\check{P}_{T_2}$ is lognormally distributed under $\mathbb{Q}_2$.

11.10 Deduce the following formula for a call option on $P(t, T_2)$ with strike K and maturity T_1:

$$C_0 = P(0, T_2)N(d_+) - KP(0, T_1)N(d_-)$$

Write down expressions for $d_\pm$.

11.11 Use put–call parity and the table of the normal distribution to find the price of the original caplet.